



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Aproximación metodológica al estudio de la variación lingüística en las interacciones digitales

De Benito Moreno, Carlota ; Estrada Arráez, Ana

Abstract: En este trabajo indagamos sobre el valor de distintos corpus del español en la red para el estudio de la variación lingüística. Para ello, comparamos los resultados del estudio de cinco fenómenos lingüísticos en dos corpus de comunicación mediada por ordenador (esTenTen, constituido por distintos tipos de textos en línea, y el corpus de tuits del proyecto Proyectando la variación lingüística de Internet). Los fenómenos lingüísticos estudiados son la elisión de la /d/ intervocálica, las formas de tratamiento en plural, el uso no referencial de ello, la pluralización de haber existencial y el uso coloquial del sufijo superlativo -érrimo. El análisis muestra que los macrocorpus compilados de forma automática sin distinguir al menos los géneros de la comunicación digital documentada pueden presentar elevadas cantidades de ruido estadístico y que, por favorecer los contextos monológicos, pueden ser problemáticos para la documentación de determinados fenómenos muy marcados (ya sea diatópica, diafásica o diastráticamente). Su mayor ventaja es la elevada cantidad de datos que ofrecen. Por el contrario, los corpus más controlados y que privilegian contextos más prominentemente dialógicos son preferibles para el estudio de las variables lingüísticas propias de la inmediatez comunicativa, a pesar de que resultan más difíciles de compilar y consultar.

DOI: <https://doi.org/10.24197/redd.1.2018.74-122>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-165921>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Originally published at:

De Benito Moreno, Carlota; Estrada Arráez, Ana (2018). Aproximación metodológica al estudio de la variación lingüística en las interacciones digitales. *Revista de Estudios del Discurso Digital (REDD)*, 1:74-122.

DOI: <https://doi.org/10.24197/redd.1.2018.74-122>

Aproximación metodológica al estudio de la variación lingüística en las interacciones digitales

A methodological approximation to the study of linguistic variation in digital interactions

CARLOTA DE BENITO MORENO

Universität Zürich, Zürichbergstrasse 9, CH-8032 Zürich

carlota.debenitomoreno@uzh.ch

ORCID: 0000-0001-9112-5471

ANA ESTRADA ARRÁEZ

Université de Liège, place du 20-Août, 7, B- 4000 Liège

ana.estrada@uliege.be

ORCID: 0000-0002-1142-439X

Recibido: 16-07-2018. Aceptado: 12-11-2018

Cómo citar: De Benito, C. & Estrada, A. (2018), “Aproximación metodológica al estudio de la variación lingüística en las interacciones digitales”, *Revista Estudios del Discurso Digital (REDD)*, 1: 74-122.

DOI: <https://doi.org/10.24197/redd.1.2018.74-122>

Resumen: En este trabajo indagamos sobre el valor de distintos corpus del español en la red para el estudio de la variación lingüística. Para ello, comparamos los resultados del estudio de cinco fenómenos lingüísticos en dos corpus de comunicación mediada por ordenador (esTenTen, constituido por distintos tipos de textos en línea, y el corpus de tuits del proyecto *Proyectando la variación lingüística de Internet*). Los fenómenos lingüísticos estudiados son la elisión de la /d/ intervocálica, las formas de tratamiento en plural, el uso no referencial de *ello*, la pluralización de *haber* existencial y el uso coloquial del sufijo superlativo *-érrimo*. El análisis muestra que los macrocorpus compilados de forma automática sin distinguir al menos los géneros de la comunicación digital documentada pueden presentar elevadas cantidades de ruido estadístico y que, por favorecer los contextos monológicos, pueden ser problemáticos para la documentación de determinados fenómenos muy marcados (ya sea diatópica, diafásica o diastráticamente). Su mayor ventaja es la elevada cantidad de datos que ofrecen. Por el contrario, los corpus más controlados y que privilegian contextos más prominentemente dialógicos son preferibles para el estudio de las variables lingüísticas propias de la inmediatez comunicativa, a pesar de que resultan más difíciles de compilar y consultar.

Palabras clave: comunicación mediada por ordenador, variación lingüística, corpus, esTenTen, Twitter

Abstract: The goal of this paper is to investigate the usefulness of different corpora of online Spanish for the study of linguistic variation. In order to do so, we compare five different linguistic phenomena in two corpus of computer mediated communication, namely, esTenTen, which consists of various types of online texts, and the Twitter corpus of the project *Proyectando la variación lingüística de Internet*. The five linguistic phenomena under study are the loss of intervocalic /d/, the use of 2nd plural personal pronouns, the non-referential use of *ello*, the plural agreement of existential *haber* and the colloquial use of the superlative suffix *-érrimo*. The analysis shows that automatically compiled macrocorpora that do not differentiate among digital genres might show a high level of statistical noise. Since they favour monological contexts, they might be problematic for the documentation of highly marked phenomena, be it diatopically, diastratically or diaphasically. The major advantage they show is the large quantities of data they offer. On the contrary, more controlled corpora that also privilege more dialogical contexts are preferable for the study of linguistic variables typical of communicative immediacy, despite the fact that they are more difficult to collect and consult.

Keywords: computer mediated communication, linguistic variation, corpora, esTenTen, Twitter

Sumario: 1. Introducción, 2. Elisión de /d/ intervocálica, 3. Formas de tratamiento, 3.1. *Ustedes* vs. *vosotros*, 3.2. Concordancia de *ustedes* con formas en 2ª pl., 4. *Ello* no referencial, 5. Pluralización de *haber* existencial, 6. El sufijo *-érrimo*, 7. Conclusiones, 8. Apéndice: el volumen de los corpus

Summary: 1. Introduction, 2. Loss of intervocalic /d/, 3. 2nd plural personal pronouns, 3.1. *Ustedes* vs. *vosotros*, 3.2. Agreement of *ustedes* with 2nd plural forms, 4. Non referential *ello*, 5. Plural agreement of existential *haber*, 6. The suffix *-érrimo*, 7. Conclusions, 8. Appendix: size of the corpora

INTRODUCCIÓN

El uso de datos provenientes de internet para la investigación lingüística está cada vez más generalizado. La enorme cantidad de información que nos proporciona la web, así como el acceso a fenómenos lingüísticos que no son fáciles de documentar en corpus más tradicionales, son dos de las grandes ventajas de esta fuente de datos. Una de las consecuencias de estas características de los datos digitales es, necesariamente, una gran variedad en cuanto a los tipos de situaciones comunicativas representadas, la cual incidirá en la utilidad de dichos datos para estudiar diferentes fenómenos lingüísticos. En este trabajo indagamos sobre el valor de los corpus de interacciones digitales para el estudio de la variación lingüística y, con este fin, realizaremos una doble comparación (cualitativa y cuantitativa) entre dos bases de datos de comunicación digital, investigando distintas clases de variación lingüística. Cada uno de los corpus representa distintos tipos de comunicación mediada por el ordenador, lo que nos permitirá averiguar sus ventajas y desventajas. Así, diferenciamos entre dos tipos de textos:

en primer lugar, aquellos predominantemente monológicos —lo cual incluye desde páginas web tradicionales, en las que predomina la función informativa y son, por tanto, intransitivas en la terminología de Cantamutto & Vela Delfa (2016), hasta datos pertenecientes a blogs, que pueden permitir una comunicación transitiva y, así, mostrar una intención algo más dialógica, especialmente en el caso de los comentarios que responden a las entradas principales—. En segundo lugar, los que son predominantemente dialógicos —como los mensajes de Whatsapp, que lo son exclusivamente, o las redes sociales como Twitter, en los que la voluntad de interacción no es condición necesaria, pero sí mayoritaria—. Para una discusión sobre el encaje de estos géneros en el continuo entre inmediatez y distancia comunicativa, cf. De Benito y Estrada (2016a).

Los corpus usados para extraer y comparar datos representativos de estos dos grupos son, en primer lugar, esTenTen, y en segundo lugar, el corpus de tuits del proyecto *Proyectando la variación lingüística de Internet*.¹ El primero, según la información que puede consultarse en su web,² está formado por más de nueve mil millones de palabras procedentes de aproximadamente veinte millones de páginas web de diversos países hispanohablantes de América y Europa. El corpus se recogió en 2011 mediante la herramienta SpiderLing,³ especialmente diseñada para la creación de corpus lingüísticos, y los resultados de las búsquedas se procesaron y limpiaron para la lematización y el etiquetado de los mismos.⁴ La asignación de la zona geográfica en este primer corpus se realizó según el dominio de la página web en cuestión, por lo que la exactitud de dicha asignación es relativamente incierta: por un lado, es muy general (solo identifica país) y, además, el uso de un dominio en concreto no nos asegura que los contenidos de ese sitio

¹ Proyecto de la Fundación BBVA concedido en 2015, con Ángel Gallego como investigador principal.
<http://www.fbbva.es/TLFU/tlfu/esp/areas/cienbas/investigacion/fichainves/index.jsp?codigo=421>

² <https://www.sketchengine.eu/estenten-spanish-corpus/#toggle-id-1>. Véase también Kilgarriff y Renau (2013).

³ <http://corpus.tools/wiki/SpiderLing>

⁴ Es debatible si este modo de recopilar datos presenta implicaciones éticas. Aunque los autores de los textos no han accedido explícitamente a la explotación científica de estos, debe tenerse en cuenta que los textos recogidos son públicos (por lo que la búsqueda adecuada en Google podría también devolverlos). Por otra parte, la información sobre los autores que ofrece el corpus es inexistente: para recuperarla deben consultarse las fuentes directas.

procedan de la zona designada. Esto es especialmente cierto en el caso de los blogs, puesto que los ejemplos encontrados dentro de la propia entrada de la bitácora digital (cuya intención interactiva es más baja) tienen una adscripción geográfica más fiable que los comentarios (los cuales, además, tienen una voluntad interactiva más evidente).

El segundo corpus contiene datos extraídos de la red social Twitter y consta de en torno a cuarenta millones de tuits en español, de los cuales hemos utilizado cerca de diecisiete millones y medio —equivalentes a más de 194 millones de palabras— pertenecientes a los países hispanohablantes de América (con la excepción de Cuba, donde la penetración de Internet es muy baja) y Europa. La extracción de los datos se hizo entre mayo y junio de 2016 a través de *R* mediante el paquete *StreamR*, el cual conecta con la API de Twitter, que ofrece una muestra aleatoria de entre el 1% y el 10% de los tuits que se están publicando a tiempo real.⁵ Este paquete permite delimitar tanto las coordenadas geográficas desde las que se emite el tuit como el idioma del mismo. Debe tenerse en cuenta, sin embargo, que solo una pequeña parte de los usuarios activa la geolocalización —algo que puede sesgar la muestra hacia un determinado tipo demográfico de usuarios⁶— y que el reconocimiento de idiomas, hecho por Twitter de forma automática, puede presentar errores. Asimismo, la geolocalización puede no ser fiable en tanto en cuanto los usuarios pueden publicar tuits desde emplazamientos distintos a su lugar de procedencia —algo también válido para las páginas web recogidas en esTenTen—.

Con el objetivo de explorar la utilidad de la comunicación digital para la investigación lingüística, compararemos la incidencia de cinco fenómenos subestándar en estos dos corpus. Los fenómenos han sido escogidos para ejemplificar tanto las distintas clases de variación (a saber: dialectal, diafásica y diastrática) como los diferentes niveles lingüísticos (fonética-fonología, morfología y sintaxis). Siguiendo a Koch & Oesterreicher (2007), partimos de la premisa de que, a mayor interacción, mayor inmediatez comunicativa y, por ello, esperamos una mayor incidencia relativa de los fenómenos más marcados diastrática y

⁵ Twitter ofrece únicamente una muestra aleatoria de los tuits públicos de esta realidad social, lo que significa que no surgen problemas éticos en la utilización de estos datos: los usuarios autorizan estos usos en las condiciones de uso de la red social.

⁶ Pavalanathan y Eisenstein (2015) señalan, entre otras cosas, que los tuits geolocalizados suelen pertenecer con más frecuencia a jóvenes de sexo femenino que a usuarios con otros perfiles, además de ser más comunes en áreas urbanas.

diafásicamente en ProVaIn (Estrada Arráez & De Benito Moreno, 2016). Asimismo, esperamos que las observaciones de índole geográfica a partir de Twitter sean más exactas que las obtenidas gracias a esTenTen. Sin embargo, las tendencias actuales en cuanto al *big data* parten de la idea de que las cantidades ingentes de datos sirven para contrarrestar el potencial ruido de las bases de datos poco controladas, por lo que la comparación de los datos de ProVaIn con los de esTenTen nos permitirá saber cuánto sacrificamos al emplear un corpus gigantesco (y de fácil acceso) en vez de un corpus de tamaño más reducido y más difícil de obtener, pero con un grado de ruido estadístico potencialmente menor. Los fenómenos explorados en esta comparación se exponen en los párrafos siguientes.

En el campo de la fonética, nos ocuparemos de la pérdida de la /d/ intervocálica (sección 2), un fenómeno que muestra una distribución dialectal irregular. Se trata de un marcador sociolingüístico, lo cual quiere decir que los hablantes son plenamente conscientes de que existe una variación asociada a la /d/ donde la elisión es menos prestigiosa, por lo que su realización cambia con el registro. También se puede observar variación diastrática, puesto que la elisión de la consonante es más común en los niveles socioculturales más bajos (Alba, 1999; Molina & Paredes, 2014). Hay que tener en cuenta que el estudio de un rasgo fonético a través de la escritura (y especialmente uno con estas características variacionales) es problemático, ya que no solamente habrá numerosos casos en los que pueda tratarse de una simple errata, sino que muchos de los hablantes que presentan el rasgo en la oralidad no lo manifestarán por escrito. Además de esto, las búsquedas deben delimitarse con mucha más precisión para evitar un número demasiado elevado de falsos positivos, ya que el contexto fonético adyacente que resulta de la elisión de la /d/ da lugar a un mayor abanico de posibilidades en cuanto al contexto general de la palabra, lo que podría devolvernos o bien elementos donde no ha habido realmente una elisión o bien casos no verdaderamente comparables entre sí (pensemos en *Bilbao*, por ejemplo).

La documentación de fenómenos morfosintácticos en textos escritos presenta menos problemas que la de los fonéticos, lo que nos permite investigar un mayor número de fenómenos. En primer lugar, investigaremos la distribución en ambos corpus de un fenómeno que muestra diferencias dialectales (a nivel nacional), pero no tanto así sociales o diafásicas (salvo en la península ibérica): el uso de *ustedes* por

vosotros (sección 3.1). Esperamos que, por no estar este fenómeno marcado diastráticamente, se encuentren profusamente documentados tanto en esTenTen como en ProVaIn. A continuación investigamos la vertiente de este fenómeno que está más marcada diafásicamente: el uso de *ustedes* con formas de 2ª persona de plural (sección 3.2). En la misma línea, en la sección 4 abordaremos otro fenómeno con una distribución dialectal muy específica y variación diafásica: el uso no referencial del pronombre *ello* (e.g. *ello hay gente*), documentado sobre todo en la región del Cibao (norte de la República Dominicana).

En tercer lugar, nos ocuparemos de un fenómeno que, además de verse afectado por la variación dialectal, presenta también comportamientos distintos según consideraciones diastráticas (y diafásicas): los usos existenciales concordados de *haber* (sección 5), tanto en primera como en tercera persona del plural (e.g. *habemos gente que no queremos ir – habían muchos niños en el parque*). Ambos usos están marcados dialectal y diastráticamente, aunque en América estos dos fenómenos pueden alcanzar niveles de lengua muy altos (Castillo Lluch & Octavio de Toledo y Huerta, 2016; Pato, 2016).

En último lugar, nos dedicamos a un fenómeno que se ve sobre todo afectado por consideraciones diafásicas, siendo típico del habla de los jóvenes y de usos lúdicos y familiares. Por lo que se nos alcanza, no ha recibido atención en la bibliografía, así que desconocemos si tiene una distribución dialectal restringida. Así pues, la comparación entre ambos corpus no tendrá únicamente interés metodológico, sino también dentro del plano de la descripción lingüística, ya que nos permitirá averiguar si este fenómeno se ve asimismo afectado por restricciones de otro tipo (dialectales o diastráticas). Se trata del uso del superlativo en *–érrimo* con un amplio número de bases adjetivales que habitualmente emplean *–ísimo* (*guapérrimo, buenérrimo*).

La comparación de estos cinco fenómenos a partir de los dos corpus ya mencionados, los cuales contienen distintos tipos de documentos digitales, puede proporcionarnos información sobre las características globales de la lengua empleada en dichos documentos gracias a la frecuencia con la que se registran fenómenos marcados diafásica, diastrática y dialectalmente. De este modo, se podrán evaluar las ventajas y desventajas de los distintos tipos de corpus de lengua para el estudio de diferentes fenómenos de variación, dependiendo de las características de estos.

1. ELISIÓN DE /d/ INTERVOCÁLICA

La pérdida de la /d/ intervocálica en interior de palabra (ejemplo 1) es un fenómeno que afecta en mayor o menor medida a todo el mundo hispanohablante, especialmente en la terminación participial *-ado* (Alba, 1999; Moya Corral & Wiedemann, 2010). Se trata de un rasgo que actúa como marcador sociolingüístico, como se adelantó en la introducción de este trabajo, ya que, excepto quizás en el caso de *-ado* (donde la aceptación de la pérdida está más o menos generalizada), la elisión se considera una variante menos prestigiosa y los hablantes son plenamente conscientes de ello, por lo que tienden a evitarla en ciertas situaciones. El español europeo (especialmente en el sur de la península ibérica) es el que muestra una mayor incidencia del fenómeno, que se manifiesta sobre todo en elementos frecuentes de significado gramatical en los que la /d/ aparece circundada por /a, o/ en cualquiera de sus combinaciones (Menéndez Pidal, 1962; Blas Arroyo, 2006; Estrada Arráez, 2017).

(1)

Pegado > [pe'ɣao]

“Dios que hostion he pegao ☹” (@14dOctubre, tuit)

Para nuestro estudio decidimos comparar la realización de la /d/ en participios de la primera conjugación (*-ado/-ao*) con los de las otras dos conjugaciones (*-ido/-ío*), los cuales deberían mostrar una tasa de elisión mucho menor (Blas Arroyo, 2006; Estrada Arráez, 2017, entre otros). La elección de estos elementos responde a varias razones. En primer lugar, reducir la búsqueda a estas formas nos permite lidiar con algunos de los problemas derivados del estudio de una variable fonética a través de la escritura, ya señalados en la introducción de este trabajo. En segundo lugar, este corpus nos da la posibilidad de comparar un contexto en el que la pérdida está aceptada de manera general con otro en el que esto no ocurre, siempre con elementos que sabemos de características similares. Finalmente, el uso de participios facilita la comparación entre zonas lingüísticas frente a elementos léxicos aislados (aunque no debemos olvidar que el uso de los tiempos compuestos cambia según el área geográfica, lo que afectará necesariamente al número absoluto de datos participiales de cada una de ellas).

Debido al enorme número de ocurrencias de los participios en *-ado*, decidimos restringir las búsquedas a las formas de la primera persona del

singular del pretérito perfecto (*he cantado*) en ambos corpus y para todas las conjugaciones, permitiendo así la comparación entre ellas.⁷ Las cifras de los datos resultantes son las siguientes: de las formas en *-ado*, encontramos 988 347 ocurrencias en esTenTen y 45 554 en ProVaIn; de las formas en *-ao*, 4 560 ocurrencias en esTenTen y 1 401 en ProVaIn; de las formas en *-ido*, 696 244 ocurrencias en esTenTen y 1426 en ProVaIn; de las formas en *-ío*, 189 ocurrencias en esTenTen y 230 en ProVaIn. En el gráfico 1 se muestra el ratio de participios por el número de palabras que compone cada corpus nacional, ordenado de mayor a menor presencia de participios en ProVaIn.⁸ Es interesante constatar que ProVaIn muestra en general una tasa de empleo del participio más elevada, lo que podría deberse a la elección de la primera persona como objetivo del estudio.

⁷ En esTenTen la búsqueda consistió en la forma *he* seguida de palabras acabadas en *ado/ao/ido/ío*. En ProVaIn, usamos la siguiente expresión regular (adaptada a los diferentes contextos): `\b[Hh][Ee]\s\\w+[Aa][Dd][Oo]\b`

⁸ El gráfico sirve también para ilustrar la distinta frecuencia de los tiempos compuestos en distintos países, con España y Centroamérica a la cabeza y los países rioplatenses en último lugar, aunque este no es nuestro objeto de estudio en este trabajo.

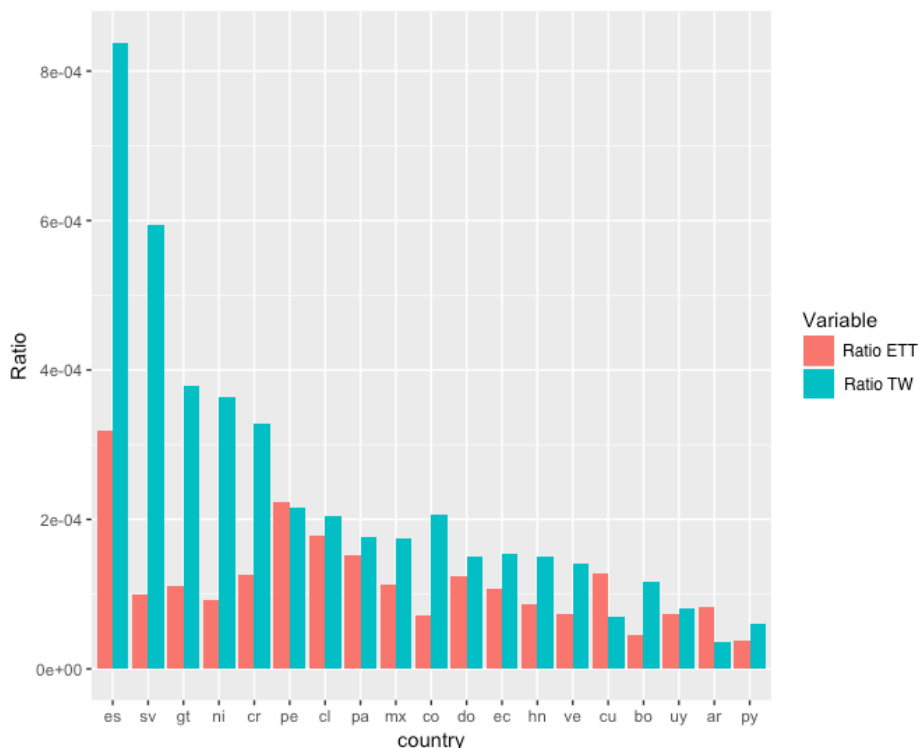


Gráfico 1. Ratio entre el número de participios y el número total de palabras según país y corpus.

En los gráficos 2 y 3 se muestra el ratio entre los casos con pérdida de los participios en *-ado* e *-ido* (respectivamente) y el total de participios por país, lo que permite observar que ProVaIn es más favorable que esTenTen a la presencia de formas con elisión de la consonante. Hacemos referencia aquí solamente a los casos en los que hay quince o más formas con elisión (marcados con una flecha), ya que solo estos permitirían realizar generalizaciones estadísticas de manera fiable. Las cifras que aparecen sobre las barras corresponden al número absoluto de elisiones de cada país y corpus.

La mayor tendencia a la elisión en ProVaIn parece mantenerse para los participios en *-ido*, aunque en este caso los datos con elisión son demasiado escasos y no permitirían un análisis estadístico fiable. En el único país en el que hay más de quince ocurrencias (España, véase también la Tabla 1), el corpus ProVaIn contiene una proporción mayor de casos con pérdida que esTenTen, lo que es un indicio claro de esta

tendencia. Esto también parece indicarnos que aquellos fenómenos con mayor aceptación exhibirán una mayor presencia en esTenTen que aquellos más estigmatizados, algo que encaja a la perfección con la mayor cercanía a la oralidad de Twitter.

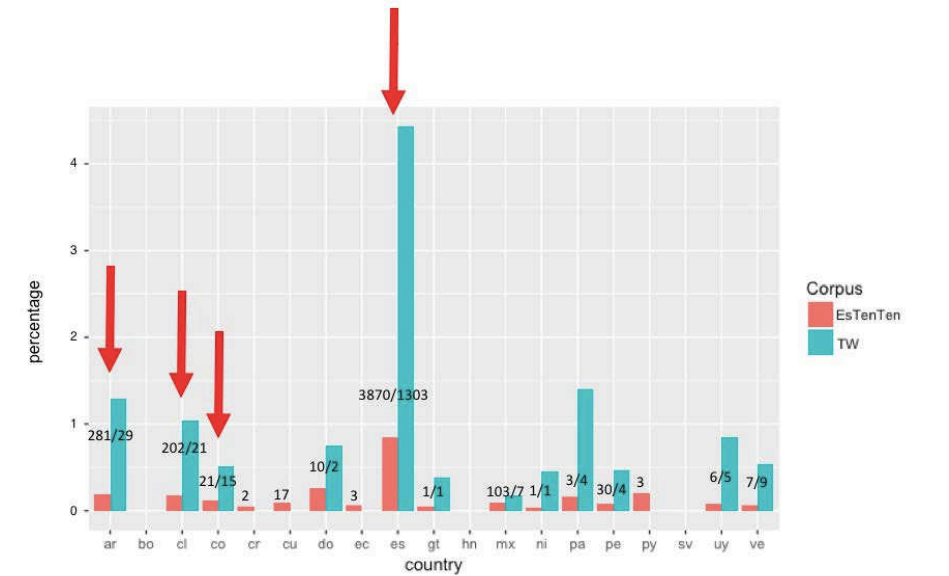


Gráfico 2. Proporción de elisión (de 0 a 100) en participios de la primera conjugación por país y corpus (con los números absolutos de elisión especificados sobre las barras).

País	ao ETT	Total ado ETT	% ETT	CI 95 % ETT	ao ProVaIn	Total ado ProVaIn	% ProVaIn	CI 95 % ProVaIn
Argentina	281	160895	0.17	0.1542454 – 0.1950507	29	2270	1.28	0.8155381 – 1.739528
Bolivia	0	1499	0	–	0	0	0	–
Chile	202	117914	0.17	0.1477068 – 0.1949158	21	2041	1.03	0.5911062 – 1.466709
Colombia	21	19037	0.11	0.06315652 – 0.1574665	15	2994	0.5	0.2480959 – 0.7539081
Costa Rica	2	4785	0.04	-0.01611869 – 0.09971325	0	0	0	–
Cuba	17	19174	0.09	0.04653331 – 0.1307902	0	0	0	–
República Dominicana	10	4017	0.25	0.09483832 – 0.4030457	2	269	0.74	-0.2830985 – 1.770087
Ecuador	3	5619	0.05	- 0.007010379 – 0.1137909	0	0	0	–

España	3870	458718	0.84	0.8171873 – 0.8701239	1303	29476	4.42	4.185884 – 4.655207
Guatemala	1	2228	0.05	-0.04306823 – 0.1328348	1	265	0.38	-0.3608673 – 1.115584
Honduras	0	560	0	–	0	0	0	–
México	103	126414	0.08	0.06574927 – 0.09720737	7	4363	0.16	0.04167979 – 0.2792003
Nicaragua	1	3692	0.03	-0.02599498 – 0.08016616	1	227	0.44	-0.4210035 – 1.302061
Panamá	3	1984	0.15	-0.01977076 – 0.3221901	4	288	1.39	0.03726299 – 2.740515
Perú	30	42109	0.07	0.04575854 – 0.09672882	4	878	0.46	0.01012979 – 0.9010319
Paraguay	3	1527	0.2	-0.0256374 – 0.4185647	0	0	0	–
El Salvador	0	2128	0	–	0	0	0	–
Uruguay	6	8782	0.07	0.01367161 – 0.1229715	5	597	0.84	0.1064821 – 1.56856
Venezuela	7	11825	0.06	0.01535613 – 0.1030371	9	1720	0.52	0.1822909 – 0.8642207

Tabla 2. Porcentajes de elisión e intervalos de confianza en participios de la primera conjugación por país y corpus.

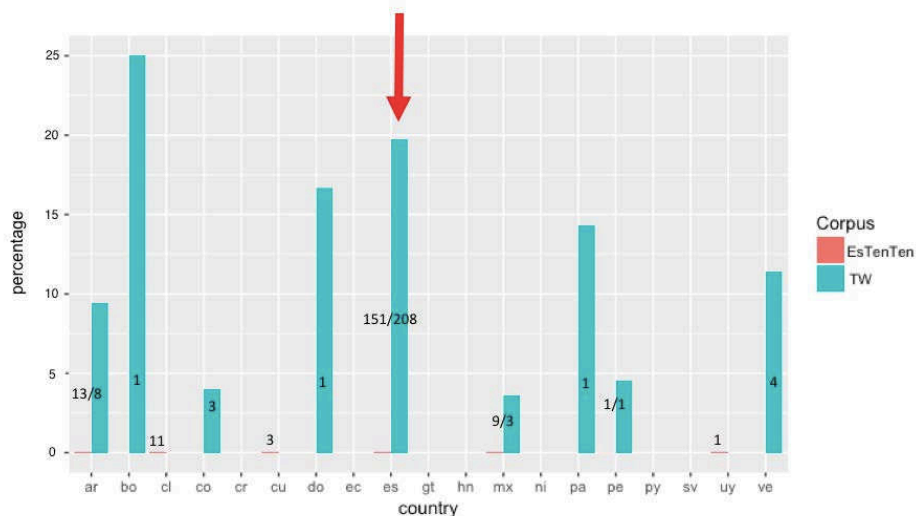


Gráfico 3. Proporción de elisión (de 0 a 100) en participios en *-ido* por país y corpus (con los números absolutos de elisión especificados sobre las barras).

País	io ETT	Total ido ETT	% ETT	CI 95% ETT	io ProVaIn	Total ido ProVaIn	% ProVaIn	CI 95% ProVaIn
Argentina	13	118357	0.01	0.005013229 – 0.01695421	8	85	9.41	3.204246 – 15.61928
Bolivia	0	0	0	–	1	4	25	-17.43524 – 67.43524
Chile	11	91119	0.01	0.004938386 – 0.01920587	0	39	0	–
Colombia	0	0	0	–	3	75	4	0.4349737 – 8.434974
Costa Rica	0	0	0	–	0	9	0	–
Cuba	3	16325	0.02	- 0.002416586 – 0.03917003	0	0	0	–
República Dominicana	0	0	0	–	1	6	16.67	-13.15378 – 46.48712
Ecuador	0	0	0	–	0	10	0	–
España	151	290330	0.05	0.04371625 – 0.06030332	208	1054	19.73	17.33158 – 22.13711
Guatemala	0	0	0	–	0	213	0	–
Honduras	0	0	0	–	0	2	0	–
México	9	94134	0.01	0.003314723 – 0.01580695	3	84	3.57	0.3971984 – 7.540056
Nicaragua	0	0	0	–	0	6	0	–
Panamá	0	0	0	–	1	7	14.29	-11.63725 – 40.20868
Perú	1	33038	0.003	- 0.002905655 – 0.00895929	1	22	4.55	4.158802 – 13.24971
Paraguay	0	0	0	–	0	4	0	–
El Salvador	0	0	0	–	0	1	0	–
Uruguay	1	6610	0.01	-0.01452121 – 0.04477839	0	10	0	–
Venezuela	0	0	0	–	4	35	11.43	0.8879832 – 21.96916

Tabla 1. Porcentajes de elisión en participios de segunda y tercera conjugación por país y corpus.

2. FORMAS DE TRATAMIENTO

Los pronombres de 2ª persona plural presentan un interesante caso de variación en el español que exhibe tanto diferencias geográficas entre distintos estándares regionales como diferencias diatópicas, diastráticas y diafásicas dentro de España. Como es bien sabido, la única forma empleada para esta persona en el español americano es *ustedes*, donde la serie *vosotros–os–vuestro*, así como las formas verbales correspondientes con lectura plural, se han perdido (salvo contextos ceremoniales, litúrgicos o muy formales, siempre muy raros, NGLE, 2009: 16.15q, 181f; Fontanella de Weinberg, 1999). En la norma peninsular general, sin embargo, la distinción entre *vosotros* y *ustedes* pervive (aunque parece que el uso está desterrando la forma *ustedes*, Morgan & Schwenter, 2016). La situación en las islas Canarias y Andalucía Occidental, sin embargo, es bastante más compleja. En Canarias predomina la forma *ustedes* en concordancia con formas etimológicamente de 3ª pl., como en español americano (Almeida & Díaz Alayón, 1988), aunque al menos hasta hace algunas décadas *vosotros* pervivía en algunas islas para el trato de confianza (Catalán, 1964, apud Fontanella de Weinberg, 1999). En La Gomera, desde luego, este uso sigue vivo, como nos ha permitido comprobar la encuesta COSER realizada en la isla en noviembre de 2018. En Andalucía occidental se documenta la forma única *ustedes*, pero con una concordancia variable (tanto en el paradigma pronominal como el verbal), que puede ir desde el único uso de las formas de 3ª pl. al uso de formas mixtas entre la 2ª y la 3ª pl. (*Ustedes se vais*). De acuerdo con Lara Bermejo (2015), esta tendencia a la desaparición de *vosotros* en Andalucía está en recesión y se ve favorecida por determinados grupos sociales (de bajo nivel educativo, rurales y de edad avanzada). Este autor, sin embargo, no encuentra que los distintos usos se vean afectados por cuestiones diafásicas. A continuación investigamos la presencia de los distintos pronombres nominativos en nuestros dos corpus (sección 3.1) y la concordancia de *ustedes* con formas de 2ª pl. (sección 3.2).

1. 1. Ustedes vs. vosotros

Para investigar la frecuencia de *ustedes* y *vosotros*, buscamos en el corpus ProVaIn todos aquellos tuits que contuvieran las formas *ustedes* y/o *vosotros/as* en mayúsculas, minúsculas o mezclando ambos tipos de

letra,⁹ pero sin considerar otras posibles variantes ortográficas. Esto devolvió un total de 41 930 casos de *ustedes* y 10 239 de *vosotros/as*. En esTenTen buscamos los lemas *ustedes* y *vosotros* separadamente para cada país,¹⁰ obteniendo un total de 1 026 851 de ocurrencias de *ustedes* y 178 915 resultados de *vosotros*.

Puesto que los pronombres de 2ª persona pueden ser en sí mismos indicativos del grado de interactividad de un corpus, por ser palabras que necesariamente se dirigen a un interlocutor (a varios, en este caso), la comparación del número de pronombres de 2ª pl. por el número de palabras de cada corpus resulta ilustrativa. Así, la suma de *ustedes* y *vosotros/as* ofrece en esTenTen un ratio de 0,1099 (el intervalo de confianza al 95 % es (0,1097-0,1101)) pronombres de 2ª pl. por cada mil palabras (1 205 766/10 968 413 182), mientras que en ProVaIn es de 0,4709 (el intervalo de confianza al 95 % es (0,4669-0,4750)) pronombres de 2ª pl. por cada mil palabras (52 166/110 776 154). Esto significa que las formas de 2ª pl. son entre 4,2 y 4,3 veces más frecuentes en nuestro corpus ProVaIn que en esTenTen, lo que da una buena idea los diferentes grados de interacción que muestran ambos corpus.

Puesto que conocemos bien el reparto de las formas en los distintos países, la comparación de los porcentajes de *vosotros/as* frente a *ustedes* nos permite averiguar, en primer lugar, el grado de fiabilidad de cada corpus (la cantidad de ruido introducido por usos seguramente no procedentes de los países en los que los documentamos) y, en segundo lugar, el grado de formalidad de cada corpus (si nos fijamos en las cifras de España). Para tener otro punto de comparación con un corpus de lengua no digital hemos añadido los resultados del CREA en el gráfico 4.

⁹ Empleamos las siguientes expresiones regulares: `"\\b[Uu][Ss][Tt][Ee][Dd][Ee][Ss]\\b"`, `"\\b[Vv][Oo][Ss][Oo][Tt][Rr][OoAa][Ss]\\b"`.

¹⁰ Búsqueda realizada el 13 de enero de 2018.

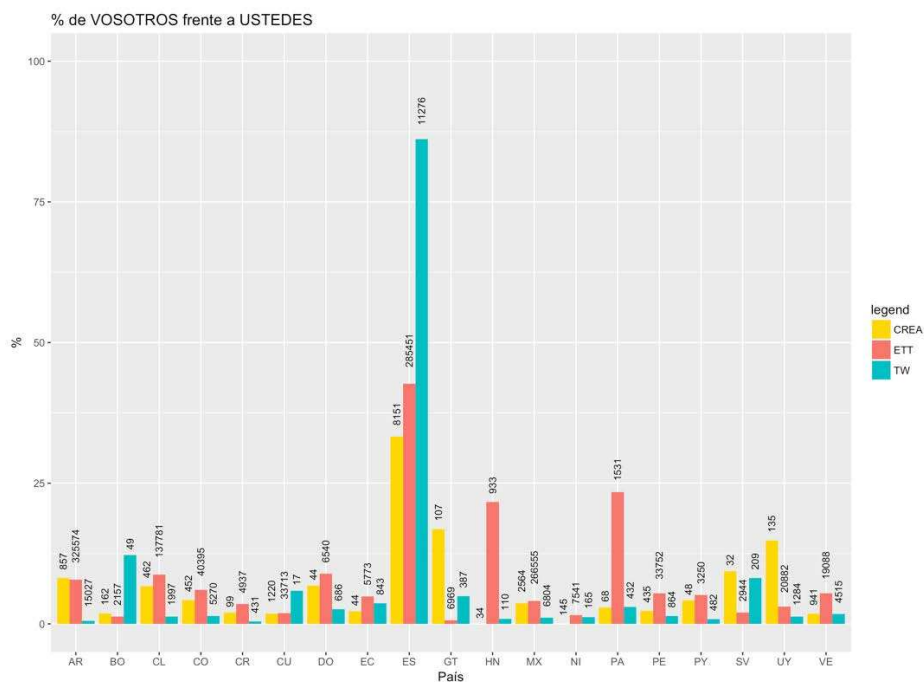


Gráfico 4. Porcentaje de *vosotros/as* frente *ustedes* por corpus y país

De la observación de dicho gráfico se desprenden varias conclusiones interesantes: en primer lugar, al menos para algunos países, esTenTen introduce una cantidad de ruido importante. Los casos de Honduras y Panamá son desde luego sorprendentes, con porcentajes de *vosotros/as* que superan el 20 %. Por otro lado, en España encontramos que esTenTen ofrece porcentajes más similares a CREA que a ProVaIn, lo que nos habla de un grado de distancia comunicativa bastante elevado de este corpus. La elevadísima frecuencia de *vosotros/as* frente a *ustedes* en el corpus ProVaIn puede interpretarse de dos maneras, que no son contradictorias, sino complementarias: la lengua documentada en Twitter es, primero, más informal y por eso se privilegian las formas de familiaridad y, segundo, el mayor grado de inmediatez comunicativa de Twitter nos ofrece un panorama más cercano al uso hablado de la lengua, en la que *ustedes* parece estar perdiéndose en España, como han demostrado Morgan y Schwenter (2016).

1. 1. Concordancia de *ustedes* con formas en 2ª pl.

Investigar la concordancia de *ustedes* resulta mucho más complejo que la mera comparación entre los dos pronombres de 2ª pl. Las búsquedas realizadas en nuestro corpus ProVaIn consistieron en extraer, en primer lugar, todos aquellos tuits que contuvieran la palabra *ustedes* (sin variación ortográfica)¹¹ y, posteriormente, eliminar todos aquellos que estuvieran precedidos inmediatamente por una preposición¹² (para reducir la búsqueda a las formas de sujeto) y, de los resultados así obtenidos, seleccionar solo aquellos tuits que contuvieran la palabra *os* o cualquier palabra acabada en *-is* o *-ís* (salvo que fuera *mis*, la más frecuente de estas en los resultados).¹³ A continuación, obtuvimos todas las palabras acabadas en *-is* o *-ís* de estos tuits y las etiquetamos manualmente, seleccionando aquellas que fueran posibles formas de 2ª persona del plural.¹⁴ Empleamos estas palabras para hacer una última

¹¹ Empleamos la expresión regular "\b[Uu][Ss][Tt][Ee][Dd][Ee][Ss]\b".

¹² Las preposiciones tenidas en cuenta fueron: *a, ante, bajo, cabe, con, contra, de, desde, en, entre, hacia, hasta, para, por, según, sin, so, sobre, tras*. Para ello utilizamos las siguientes expresiones regulares: "\b([aA][Aa][Nn][Tt][Ee][Bb][Aa][Jj][Oo][Cc][Aa][Bb][Ee][Cc][Oo][Nn][Cc][Oo][Nn][Tt][Rr][Aa])\b\s[Uu][Ss][Tt][Ee][Dd][Ee][Ss]\b"; "\b([Dd][Ee][Dd][Ee][Ss][Dd][Ee][Ee][Nn][Ee][Nn][Tt][Rr][Ee][Hh][Aa][Cc][Ii][Aa][Hh][Aa][Ss][Tt][Aa])\b\s[Uu][Ss][Tt][Ee][Dd][Ee][Ss]\b"; "\b([Pp][Aa][Rr][Aa][Pp][Oo][Rr][Ss][Ee][Gg][Uu][Úú][Nn][Ss][Ii][Nn][Ss][Oo][Ss][Oo][Bb][Rr][Ee][Tt][Rr][Aa][Ss])\b\s[Uu][Ss][Tt][Ee][Dd][Ee][Ss]\b".

Nótese que esto no ha excluido todo los casos en que *ustedes* está precedido de preposición: tenemos casos como *pa ustedes* o *para todos ustedes* que se eliminaron de forma manual.

¹³ Para ello, seleccionamos los tuits que encajaban con la siguiente expresión regular "(\\b[Oo][Ss]\\b[Iíí][Ss]\\b)" y que excluyeran "\\b[Mm][Ii][Ss]\\b".

¹⁴ Las palabras seleccionadas fueron: *sois, decís, decis, estáis, tenéis, queréis, habéis, sabéis, vais, podeis, sabeis, teneis, dais, tenis, venis, entendéis, estais, habeis, hagáis, mandáis, merecéis, podéis, poneis, veis, bebis, creéis, deciis, dejais, ganais, haceis, hacéis, pedís, ponéis, salis, salís, sentís, tendréis, venís, vivis, vivís, abriis, acertais, alegrareis, animais, aparecéis, apoyéis, aprendéis, aprobaréis, arrepentis, asesoráis, avisáis, ayudais, bais, caéis, cagáis, cargáis, celebrabais, celebrais, cobráis, coláis, comprenderéis, condenais, condenáis, convirtiis, creiais, criticáis, cumplís, debéis, decidis, decidís, decidisteis, deis, dejáis, desaparecisteis, descoméis, digais, diréis, disfrutareis.qe, divirtais, dormís, echáis, elegís, ensuciais, escribis, escribís, estabais, estareis, estaréis, exponeis, falléis, faveeis, ganáis, ganareis, ganasteis, habláis, hacéis, hundiríais, insultais, insultasteis, interactueis, jugais, kerais, liais, lloréis, malmeteis.ojo, mamáis, mancháis, matáis, mentis,*

criba, extrayendo los tuits que contuvieran estas y *os*, lo que nos dio un resultado de 205 ejemplos que fueron analizados manualmente, de los que descartamos 45 por ser casos en los que *ustedes* no era sujeto, no eran casos de concordancia de 2ª pl. (sino de voseo en el mismo tuit, por ejemplo), erratas, etc.¹⁵ De esta manera nos quedamos con lo que resulta en 160 casos en los que *ustedes* aparece con formas de la 2ª pl. en el mismo tuit. Esta coaparición puede ser de varios tipos: 1) *ustedes* y las formas de 2ª pl. aparecen en la misma oración (2a); 2) *ustedes* y las formas de 2ª pl. aparecen en oraciones distintas y la oración que contiene *ustedes* no contiene formas que denoten concordancia personal (2b), 3) *ustedes* concuerda en su oración con formas de 3ª pl., pero en otras oraciones colindantes el interlocutor aparece referido con formas de 2ª pl. (2c). Nos referimos a estos tres tipos como discordancia “cercana”, “lejana” y concordancia “doble” respectivamente.

(2)

- a) A ver si los que la estáis viendo como un objeto sois ustedes y no es ella la que se presenta como tal, eh. No sé, podría ser. Quizá. (Granada, España)
- b) manda os vacila y ustedes cayendo en serio qué os pasa JAJAJAJAJA (Las Palmas de Gran Canaria, España)
- c) Ustedes no tienen esa Inteligencia y la claman. Todo se os caerá si no cumplís y dejais las primeras líneas que os llevaron a inmiscuiros (Sevilla, España)

Como podemos ver en la tabla 3, la mayoría de los ejemplos (83,8 %) encontrados muestran ejemplos de discordancia cercana:

metéis, moveis, necesiteis, ois, pagáis, parais, parecéis, pasáis, pedis, podríais, pongáis, proponeis, provocais, quemáis, querais, queráis, querais, realizáis, recibiríais, referís, reís, remontáis, riáis, sabís, salíais, seguís, seréis, sonéis, subís, superasteis, suscribis, teniais, teseis, tocais, trabajareis, traéis, twitteáis, unis, vayáis, vengáis, vereis, vestis, volvierais, votais, votáis

¹⁵ Para la identificación de las potencialmente formas voseantes hemos sido conservadoras (en todos los países), por lo que muchas veces el parámetro que nos ha ayudado a decidírnos ha sido la presencia del pronombre *os*.

	Ustedes + 2ª pl.
Discordancia cercana	134 (83,8 %)
Discordancia lejana	9 (5,6 %)
Concordancia doble	17 (10,6 %)
Total	160 (100%)

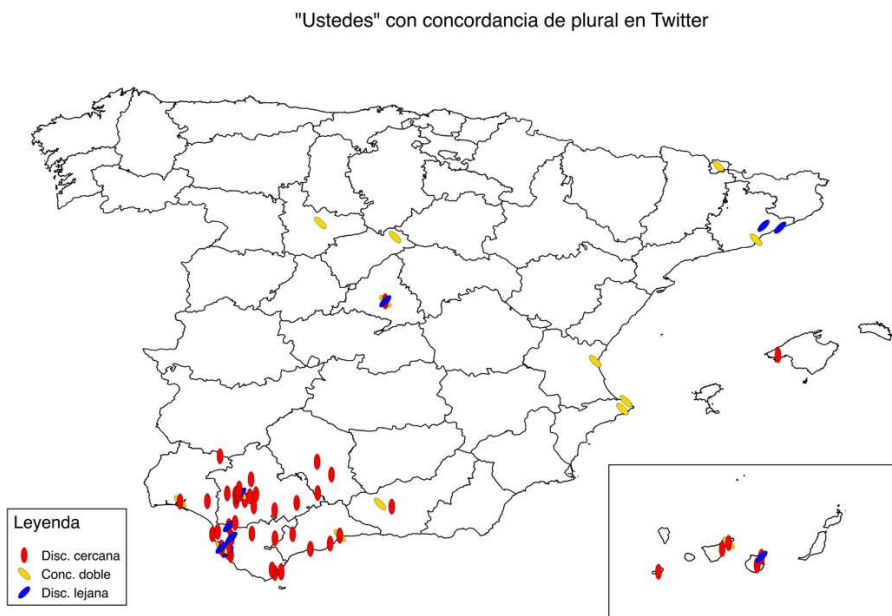
Tabla 3. Tipo de concordancia mixta con *ustedes* en el corpus ProVaIn

En la tabla 4 observamos como la gran mayoría de estos ejemplos (95,1%) proceden, según lo esperado, de España:

País	Ustedes + 2ª pl.
Andorra	1 (0,6 %)
Chile	1 (0,6 %)
Colombia	1 (0,6 %)
España	152 (95 %)
Gibraltar (RU)	1 (0,6 %)
México	2 (1,2 %)
Uruguay	1 (0,6 %)
Venezuela	1 (0,6 %)
Total	160 (100 %)

Tabla 4. Concordancia mixta con *ustedes* por país en el corpus ProVaIn

De acuerdo con la conocida ausencia de *vosotros* y de sus formas del plural en América, parece razonable deducir que los ejemplos localizados allí se deben a ejemplos de españoles que, por el motivo que sea, tuitearon desde allí, aunque no debemos olvidar que dicha forma se enseña o enseñaba hasta hace poco en las escuelas americanas. El mapa 1 muestra cómo, dentro de España, la mayoría de los puntos están concentrados en Andalucía occidental, como era de esperar. La cuestión, claro está, es si podemos atribuir todos los casos restantes a ejemplos de movilidad de los tuiteros o, especialmente en el caso de las islas Canarias, donde es habitual el uso del pronombre *ustedes* y —al menos para la mayoría de los hablantes— la ausencia de *vosotros*, se trata de ejemplos “genuinos”, es decir, de hablantes canarios que presentan la concordancia mixta en su sistema gramatical (especialmente por el elevado número de casos de discordancia cercana que presenta).



Mapa 1. Distribución de la concordancia de *ustedes* con formas de 2ª pl. en España en el corpus ProVaIn

Las búsquedas que realizamos en el corpus esTenTen consistieron en buscar la palabra *ustedes* excluyendo que estuviera precedida inmediatamente por una preposición, lo que produjo un total de 523 689 resultados en el subcorpus americano. Como la herramienta permite exportar hasta un máximo de 100 000 resultados, solo pudimos obtener un total de 387 008 ejemplos, a pesar de haber maximizado el resultado al separar las búsquedas por países (en México y Argentina había más de 100 000 casos, por lo que no pudimos extraer todos los casos).¹⁶ Sobre este conjunto aplicamos las mismas cribas que ya habíamos aplicado en el corpus ProVaIn, es decir, seleccionar aquellos casos que además contuvieran la palabra *os* o palabras acabadas en *-is/-ís* (excluyendo *mis*) y, posteriormente, seleccionar aquellos tuits que contuvieran *os* o la lista de formas que potencialmente pudieran ser verbos en 2ª pl. que anotamos manualmente.¹⁷ De esta manera nos quedamos con 596 ejemplos que

¹⁶ Lamentablemente, esTenTen no permite extraer una muestra aleatoria de dichos resultados, sino que ofrece los 100 000 primeros.

¹⁷ Dichas formas fueron: *decís, tenis, decís, soís, pelolais, habéis, sentís, podéis, solís, venís, monsváis, queréis, sentís, existís, pedís, sabeis, teneis, venís, creeis, escribis*,

analizamos a mano. En el subcorpus español, la misma búsqueda devolvió 133 574 resultados, de los que solo pudimos exportar 100 000. Tras aplicar las mismas cribas descritas anteriormente obtuvimos 686 ejemplos para analizar manualmente.¹⁸

estáis, habeis, referis, seguís, vivís, anais, beis, conoceréis, estais, podeis, salís, seréis, tenís, vais, vivís, creáis, discrimináis, esteis, fuisteis, habréis, haceis, hacéis, hacís, hagáis, mentís, pedís, perdís, preocupeis, queris, referís, rendís, sabís, seguís, seris, servís, temáis, tenéis, vereis, viviréis, acusáis, amáis, andaréis, apretáis, ariais, arrepentís, ayudais, bebáis, bebis, bibis, brandeis, brindáis, buscais, calláis, cambiáis, cantáis, coatís, coincidís, coméis, compráis, comprendéis, comprimís, condenáis, condeneis, confundís, conoceis, conocereis, conozcáis, conseguís, contenís, convertís, creais, creis, criticáis, dais, dcis, debeis, debis, decidís, defendéis, deis, dejeis, denunciáis, describís, descubris, desesperéis, dijeseis, direis, dirigís, discutís, dormís, dormís, empezáis, enredáis, ensuciéis, escribís, escuchais, escucháis, esperáis, evocáis, exigís, existís, fuerais, gustéis, habáis, habreis, incluís, inferis, iréis, iris, jodáis, juzguéis, kereis, leais, llameis, llevais, metáis, morís, ondulaís, opinais, opináis, paralizáis, partís, pensareis, penséis, permitís, pertenecéis, pintáis, podís, podís, podeis, podríais, portáis, precupeís, preferís, presteis, pretendéis, pusieseis, quereis, recibís, recibisteis, reflexioneis, reneguéis, representáis, sabéis, sabis, salís, servis, sigais, sóis, solucionareis, subís, sufrireis, sugeris, tendreis, teneís, tengáis, traeréis, transmitais, utilizais, vayais, vayáis.

¹⁸ Las palabras que podían constituir formas verbales de 2pl fueron en este caso: *sois, habeis, habéis, teneis, vais, estais, podeis, sabeis, tenéis, creéis, estáis, haceis, quereis, sabéis, tenis, veis, decis, opinais, podéis, dais, llevais, pensais, poneis, pasais, llamais, queréis, conoceis, entendeis, escribís, esteis, hacéis, jugais, meteis, quedais, sepais, vereis, abeis, aceis, aprendais, cobrais, conocéis, decís, dedicáis, dejéis, estabais, estábais, estaréis, estéis, ganais, habreis, hagais, hicisteis, kereis, kisierais, llamáis, llegais, mirais, pensareis, permitís, podáis, querais, queráis, quisierais, recomendais, reis, seais, sereis, teníais, tuvisteis, veáis, vendeis, vengais, abusais, acabasteis, acertareis, acordais, acordáis, acostáis, acostumbrais, alegrábais, aplaudíais, aportais, aprendereis, aprobais, aprobáis, aprovecharíais.me, apuntais, areis, aris, aspirais, aveis, ayudais, basáis, botais, cambeis, cambieis, canséis, cantáis, comentábais, comentais, comparais, compareis, comprendéis, conformais, conseguís, consigais, construisteis, contais.en, contaminais, continuáis, contratáis, contrateis, creais, creéis, criticaís, cubriais, daís, dáis, dariáis, debatís, debéis, deberíais, deciais, decíais, decidáis, dedicais, deis, dejais, dejareis, dejasteis, desalentéis, desconoceis, desíais, desprendéis, digais, dignábais, direis, diríais, disfrutais, dormís, eliminasteis, enpezais, enteráis, escuchais, estareis, estaríais, estuvisteis, fallais, fiáis, fichábais, fijais, fuisteis, furnis, ganareis, ganaréis, ganasteis, ganeis, goleasteis, habiaís, hablabais, habláis, hableis, habréis, hagáis, haréis, haveis, hayáis, hubierais, ibais, imaginabais, imaginéis, informarais, intentais, ireis, iríais, jodeis, jugasteis, kerais, leeis, llebais, llegueis, llevasteis, lleveis, lograreis, mandais, meais, mentiís, merecéis, metíais, mojais, multais, noteis, opináis, opinareis, opineis, opinéis, ovulais, paguéis, pasareis, pasateis, pensabais, penséis, perdais, perdeis, pilléis, piseis, podiais,*

El análisis manual de los resultados ofreció un total de 519 resultados de concordancia de *ustedes* en 2ª pl. en total, 404 de los cuales se atestiguan en el corpus de España, frente a los 115 resultados del americano; una diferencia esperable teniendo en cuenta que el fenómeno se documenta principalmente en Andalucía occidental. En la tabla 5 pueden observarse las proporciones de los distintos tipos de concordancia en cada subcorpus.

Ustedes + 2ª pl.	Discordancia cercana	Discordancia lejana	Concordancia doble	Total
España	363 (89,9 %)	12 (3 %)	29 (7,2 %)	404 (100 %)
América	77 (67 %)	4 (3,5 %)	34 (29,6 %)	115 (100 %)
España + América	440 (84,8 %)	16 (3,1 %)	63 (12,1 %)	519 (100 %)

Tabla 5. Tipo de concordancia mixta con *ustedes* en esTenTen por subcorpus

Si bien los datos americanos resultan en principio “sospechosos”, pues no esperamos encontrar allí el fenómeno, resulta muy interesante observar la elevada frecuencia de los casos de concordancia doble, que podrían sugerir que, al menos en la escritura, la combinación entre ambas formas no es desconocida en las variedades americanas. Esta combinación, sin embargo, tendría características distintas a las que tiene en la Península, pues no habría llegado a crear un paradigma mixto propiamente dicho, sino que habría dado lugar al uso de las formas de *ustedes* y las de *vosotros/as* en el mismo discurso, pero no en la misma oración. Los países americanos que más ejemplos aportan son Argentina (38 ejemplos; 0,011 casos por millón de palabras), México (29 ejemplos; 0,015 casos por millón de palabras) y Chile (14 ejemplos, 0,012 casos por millón de palabras), lo que está en consonancia con el tamaño de sendos subcorpus y, por lo tanto, no nos habla de una concentración acusada de estas construcciones en un área americana concreta.

podréis, podríais, pongáis, poseéis, preferis, prenis, presenteis, pretendéis, quedareis, quejais, queriais, queráis, reconoceis, recordais, recordasteis, retireis, revalorizais, sabiais, sabreis, sabréis, salgais, salis, saltais, seáis, sepáis, sereis, servis, sigais, sigáis, siois, spis, tais, tendereis, tengais, tocais, trabajais, traisteis, trajisteis, trastearais, tratáis, usareis, utilizais, vasilais, veís, vendreis, venis, venís, veréis, viarais, viérais, vinis, volvierais, votáis, voteis, yegais.

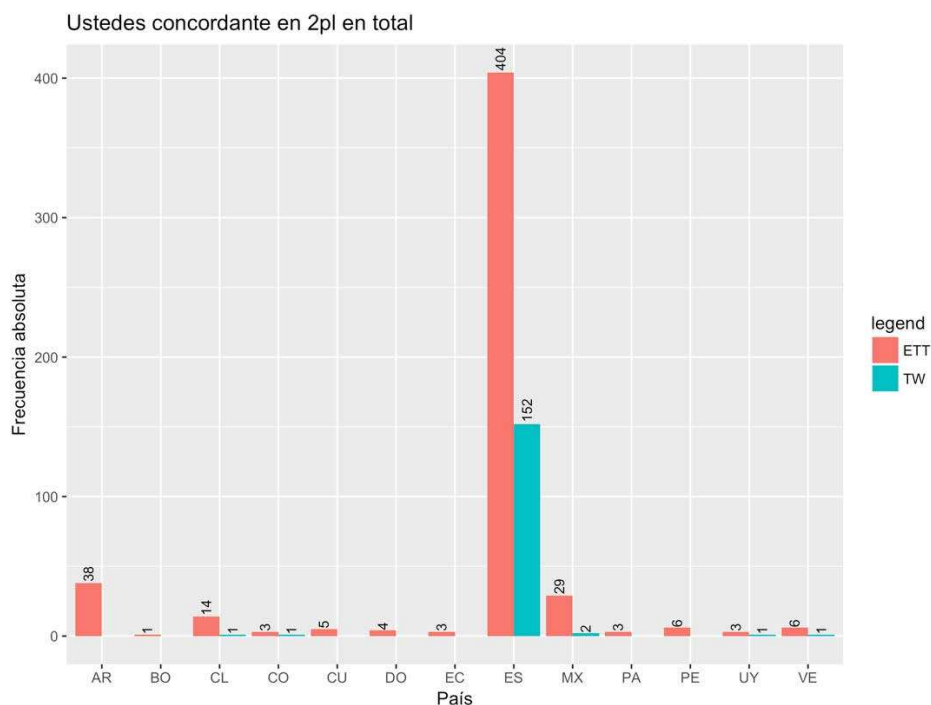


Gráfico 5. Frecuencia absoluta de *ustedes* con concordancia mixta por corpus y país

La comparación gráfica entre los resultados del corpus ProVaIn con los de esTenTen ilustra con claridad que la ventaja de esTenTen es su gran tamaño, que permite documentar más de tres veces más ejemplos que nuestro pequeño corpus de tuits (gráfico 5), mientras que la superioridad de ocurrencias por millón en ProVaIn (gráfico 6) nos habla de la mayor proximidad de este corpus a la inmediatez comunicativa y sugiere que, en un corpus de tuits de igual tamaño a esTenTen podríamos obtener un número sustancialmente más elevado de ejemplos que permitiera una minuciosa investigación.

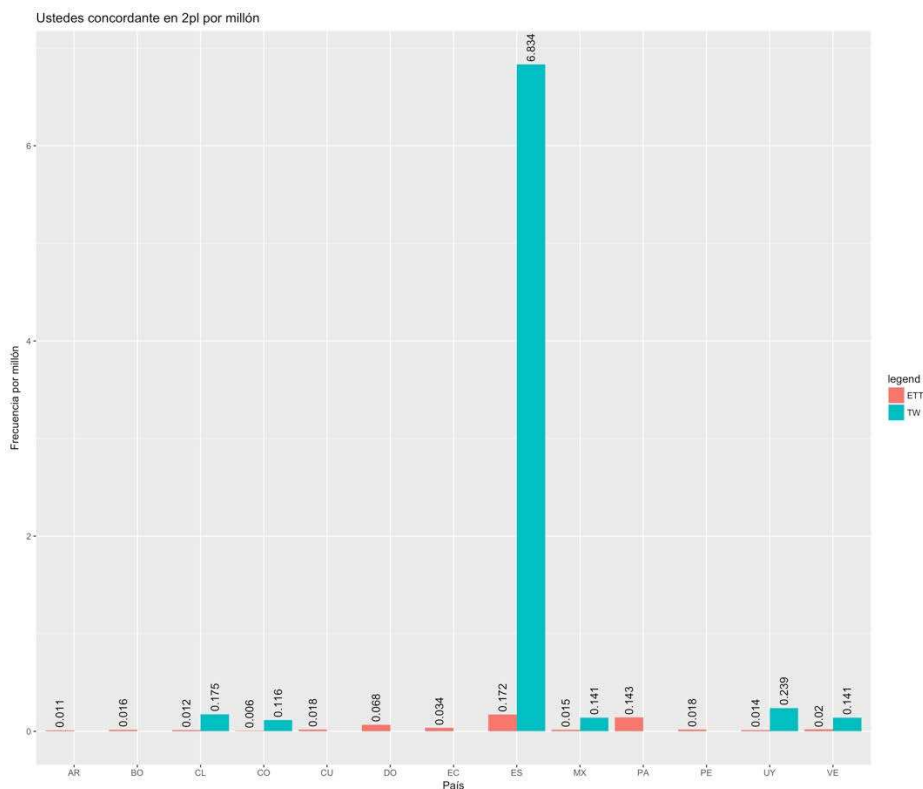


Gráfico 6. Frecuencia por millón de palabras de *ustedes* con concordancia mixta por corpus y país

3. *ELLO* NO REFERENCIAL

En los registros coloquiales de la República Dominicana el pronombre *ello* tiene un uso que recuerda a los sujetos expletivos de las lenguas sin sujeto nulo. Estructuras tales como *it rains / il pleut / es regnet* en inglés, francés y alemán parecen tener un correlato en la oración *ello llueve*, algo que se ha sugerido en algunas ocasiones. Sin embargo, y aunque ciertos autores consideran que este pronombre tiene la función de sujeto expletivo (Muñoz Pérez, 2014; Toribio, 2000), otros lo presentan como un elemento con un valor pragmático más que sintáctico (Hinzelin & Kaiser, 2007; Hinzelin, 2010; Corr, 2016). Las estructuras en las que puede aparecer este pronombre con un uso no referencial son las siguientes (Hinzelin & Kaiser, 2007; Henríquez Ureña, 1939):

- En estructuras impersonales
 - (3) *Ello hay gente*
 - (4) *Ello llueve*
- Con verbos inacusativos o con verbos usados en voz pasiva
 - (5) *Ello llegan guaguas*
 - (6) *Ello se vende arroz*
- Construcciones copulativas o con verbos de apoyo
 - (7) *Ello es fácil*
- Con sentido concesivo o evasivo (en cuyo caso puede aparecer incluso separado por una coma del resto de la oración) o, aislado, para expresar incertidumbre
 - (8) *Ello veremos*
 - (9) *Ello, yo ya le diré*
 - (10) *—¿Quieres bailar? —Ello...*
- En construcciones sin verbo, para expresar énfasis
 - (11) *Ello no obstante...*
- Delante de *sí* o *no*, para reforzar la afirmación o la negación
 - (12) *Ello no, es que las cosas son así*

Según Hinzelin y Kaiser (2007), el hecho de que el uso del pronombre no se restrinja a contextos impersonales y la falta de obligatoriedad del mismo en estos últimos apoyaría el valor meramente pragmático de *ello*. El pronombre no tendría entonces la función de sujeto en ningún caso, sino que nos daría información sobre la actitud del hablante hacia el fragmento del discurso al que se esté refiriendo. Corr (2016), por su parte, toma una posición intermedia y defiende que el pronombre tiene tanto propiedades pragmáticas como expletivas, ya que, aunque no es en ningún caso obligatorio y parece ser utilizado para reforzar el valor asertivo de la oración, su uso también depende de ciertas propiedades gramaticales. Toribio (2000) toma una posición más extrema y considera que en la República Dominicana los pronombres sujeto pueden en general incluirse libremente en una oración sin que ello implique un valor pragmático añadido, por lo que estima que se trata simplemente de pronombres expletivos. La misma interpretación aporta Muñoz, que propone que *ello* se trata de “a null subject expletive that has failed to be unpronounced” (2014: 159).

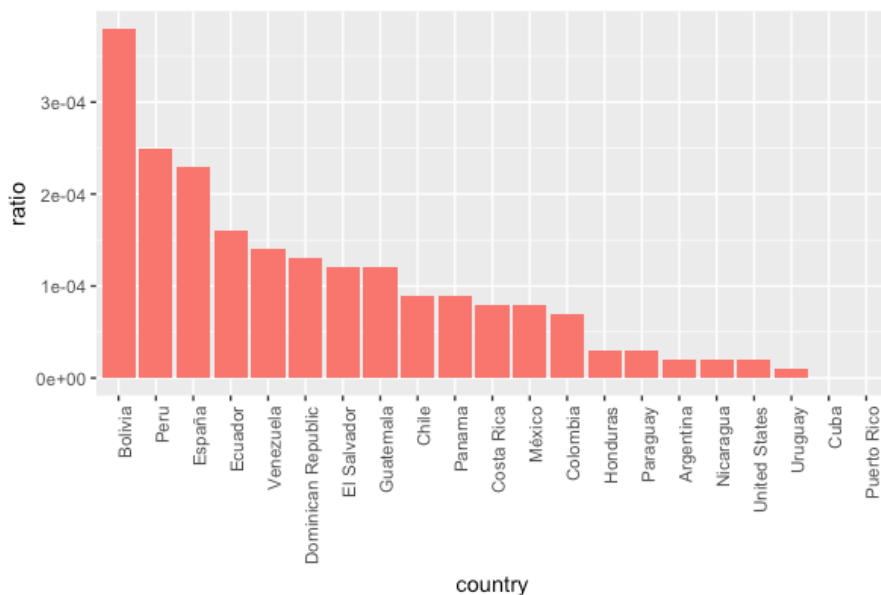


Gráfico 7. Ratio entre el número de ocurrencias de *ello* y número de tuits, por país.

El uso no referencial del pronombre tiene una distribución geográfica bastante limitada y, como ya hemos señalado, se trata de un rasgo que aparece fundamentalmente en el registro coloquial, por lo que documentar el fenómeno en los corpus tradicionales es harto complicado. La búsqueda de *ello* en el corpus ProVaIn¹⁹ devolvió 1 532 ocurrencias en todas las zonas geográficas, de las cuales solo 18 corresponden a la República Dominicana.²⁰ En el gráfico 7 se muestra el ratio entre el número de ocurrencias del pronombre y el número de tuits por país.²¹

Llama la atención el hecho de que, a pesar de contar con tan solo 18 ocurrencias del pronombre en la zona que nos interesa, nada menos que 5 de ellas corresponden a un uso no referencial (los ejemplos de (13)). Cabe destacar que el ejemplo (13c) parece ser un uso metalingüístico: este tuit es la respuesta a otro (imagen 1) que podría entenderse como una crítica a la pluralización de la primera persona del verbo *haber*. El

¹⁹ La expresión regular utilizada para esta búsqueda fue `"\b[Ee][Ll][Ll][Oo]\b\s\w+([AaEeOoYy])\xc1\x3\x1\x3\x3\x3\b"`.

²⁰ El total de palabras correspondientes a este país dentro del corpus es de 1 847 325.

²¹ El elevado número de ocurrencias del pronombre *ello* en algunos de los países que aparecen en el gráfico (especialmente en Bolivia) requiere un estudio en mayor profundidad para saber qué es exactamente lo que provoca esta diferencia con respecto al resto de áreas geográficas.

ejemplo (13c) se podría tomar en ese caso como una muestra de la estigmatización del fenómeno.

(13)

- a) **Ello** hay @Ciplaci <https://t.co/Nf5eLHD0NG>
- b) Dichosos y bienaventurados aquellos que podéis echar vainas Ella puede porque "ELLO HAY AHÍ" ... <https://t.co/SjkMkZMmTC>
- c) **Ello** hay gente <https://t.co/52FYnL1Sv3>
- d) **Ello** hay gente que se levanta día a día con el firme propósito de joder al otro y ganarse la mala voluntad de todos. #PoNá
- e) **Ello** hay orgasmo Precoz?



Imagen 1. Respuesta a (13c)

Si bien es cierto que el porcentaje de usos no referenciales con respecto al total de ocurrencias de *ello* en República Dominicana es muy elevado (en torno al 27 %), lo cual es de por sí bastante interesante, la escasez de datos nos impide sacar cualquier tipo de conclusión sobre el fenómeno. Para obtener una mejor idea de la frecuencia de estos usos analizamos algunos de los contextos en los que se documenta el *ello* no referencial. En la tabla 6 pueden observarse las ocurrencias de las estructuras sin pronombre que hemos explorado (*hay*, *llueve*, *es fácil*, *es difícil*) en comparación con las ocurrencias de esas mismas estructuras acompañadas de *ello*. Dichas estructuras son relativamente numerosas (a excepción del caso de *llueve*), aunque es la tercera más frecuente de entre ellas la única que coocurre con el pronombre no referencial. Esto está en consonancia con lo dicho en la bibliografía, donde se señala que la estructura con *haber* es, de las aquí mostradas, la que presenta el pronombre con más frecuencia (Hinzelin, 2010; Corr, 2016).

Forma	Con <i>ello</i>	Sin <i>ello</i>
Hay	5	3089
Llueve	0	99
Es fácil	0	4994
Es difícil	0	5237

Tabla 6. ProVaIn (subcorpus de la República Dominicana, 149 845 palabras).

Si comparamos estas cifras con las de nuestro otro corpus principal (tabla 7), podemos comprobar que, a pesar de todo, ProVaIn parece proporcionalmente una mejor fuente de datos para este fenómeno (sin perder nunca de vista que, al ser el número de casos tan bajo, estas observaciones no permiten comparaciones estadísticas fiables), ya que este corpus ofrece un único ejemplo (en (14)), lo que significa una menor proporción de ocurrencias en números tanto absolutos como relativos.

Forma	Con <i>ello</i>	Sin <i>ello</i>
Ello hay	1	59 907
Ello llueve	0	340
Ello sí/no	0	391 779
Ello es fácil/difícil	0	1 516

Tabla 7. EsTenTen (sección de español americano, subcorpus de la República Dominicana, 58 millones de formas).

(14) Uno de los obreros alzaba su cuello para gritarle: “No se preocupe doña, que **ello hay** más hombres que quieren ser periodistas”.

Con el objetivo de contextualizar mejor este contraste entre ambos corpus consultamos también la sección web el Corpus del Español (cuya composición es muy similar a la de esTenTen, en tanto en cuanto también reúne documentos de páginas web de distintos tipos extraídos de forma automática) y el CORPES XXI, que no es un corpus de comunicación mediada por ordenador. La situación que observamos en esTenTen se repite con los datos de estos dos corpus, como se ve en las tablas 8 y 9 y los ejemplos (15) y (16). Tres de los cuatro ejemplos encontrados en el Corpus del Español son referencias metalingüísticas a la construcción que nos interesa, por lo que no sirven como ejemplos de uso, mientras que lo único que permite confirmar el único caso que aporta el CORPES XXI es la procedencia dominicana del fenómeno.

Forma	Con <i>ello</i>	Sin <i>ello</i>
Ello hay	4	51 676
Ello llueve	0	224
Ello sí / no	0	382 983
Ello es fácil / difícil	0	2 122

Tabla 8. Corpus del Español (sección web, subcorpus de la República Dominicana, 19 millones de palabras)

(15) Corpus del Español

- a) “por que hay gente,, **ELLO HAY**,, jejeje, como dicen por ahí” (encierrogris.blogspot.com)
- b) “(ello hay jajajaja conozco unas cuántas que dicen así ayyyyy Dios Míiiiioooo!)” (encierrogris.blogspot.com)
- c) “Nota: cual es el problema con el **ello hay**, como es que se dice???” (encierrogris.blogspot.com)
- d) “**Ello hay** personas (Si, está bien empleado el término. Ya se puede decir ‘Ello hay’), que por más que se esfuercen en quedar bien, en caer bien, en hacer las cosas bien, siempre hay algo que les lleva la contraria a sus acciones.” (laveraddominicana.wordpress.com)

Forma	Con <i>ello</i>	Sin <i>ello</i>
Ello hay	1	154 063
Ello llueve	0	838
Ello sí / no	0	1 761 587
Ello es fácil / difícil	0	6 819

Tabla 9. CORPES XXI (sección de español americano, 147 millones de formas).

(16) CORPES XXI

“Repíte el 66, mirándome. Se da cuenta de que no ha logrado meterme terror. Se le apagan los ojos. «**Ello hay** gente que no cree», deja en el aire”. (*Palomos*, Pedro Antonio Valdez)

4. PLURALIZACIÓN DE *HABER* EXISTENCIAL

En diversos países de habla hispana²² se han atestiguado formas pluralizadas del verbo *haber* como las que se muestran en los ejemplos (17). En ellos, el verbo de la oración (impersonal en la lengua estándar) concuerda con su complemento, un fenómeno que se ha relacionado con

²² Se documenta el fenómeno en Argentina, Chile, Ecuador, España, Colombia, México, Perú, Puerto Rico, República Dominicana, Venezuela, Bolivia, Paraguay, Uruguay, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua y Cuba (Bentivoglio & Sedano, 1996; Fontanella de Weinberg, 1992; Pato, 2016; Quintanilla-Aguilar, 2009; Castillo-Trelles, 2007, entre otros). Según la NGLE (2009) hoy en día el fenómeno tiene más intensidad en América.

la rareza propia de una estructura impersonal transitiva y el contraste con otros verbos personales de significado parecido, como por ejemplo *existir* (NGLE, 2009: 41.6c). Son varios los autores que indican que la variación relacionada con este fenómeno ya no es diastrática, sino diafásica, aunque la valoración social de la concordancia y la penetración del fenómeno en los diferentes registros varía enormemente según la zona geográfica (Quintanilla-Aguilar, 2009; Claes, 2014 *apud* Pato, 2016). En Argentina, por ejemplo, la pluralización está muy extendida a todos los registros sociales, especialmente para la tercera persona, al igual que en Chile, mientras que en México, por ejemplo, tiene poca extensión (Fontanella de Weinberg, 1992; Castillo-Trelles, 2007; Bentivoglio & Sedano, 1996).

(17)

- a) En mis épocas solo **habían** 150 pokemones
- b) Grabé 3 minutos y **hubieron** 20 o más truenos
- c) Solo **habemos** dos Ricardos en ese piso
- d) En el centro también **hayn** cafés (Medina Cosas [Méx. 1990], *apud* DPD).

Como se señala en la NGLE (2009: 41.6d), “el uso concordado de *haber* no se extiende por igual a todas las formas del paradigma, ya que, incluso en las áreas lingüísticas en las que la concordancia está más generalizada se acepta con mayor naturalidad *Habían dificultades* que *Hubieron dificultades*”. La forma pluralizada del presente (*hayn*) es la más rara —debido a sus peculiaridades morfológicas—, aunque se ha documentado en México (DPD), Cuba (Claes, 2016), Fuerteventura (Samper Padilla & Hernández Cabrera, 2012) y El Hierro (Pérez Martín, 2007).²³ Por otro lado, las formas en 1ª pl. (*habemos*), que muestran la inclusión del hablante, están mucho más estigmatizadas que las formas de 3ª persona (NGLE, 2009: 41.6d, Castillo Lluch & Octavio de Toledo, 2016).

Las búsquedas realizadas para el estudio de la concordancia en tercera persona corresponden a las formas *habían* y *hubieron* en diferentes variantes gráficas (con *b/v*, tildadas o sin tildar y con

²³ También la hemos documentado con cierta profusión en la campaña de encuesta COSER (Fernández-Ordóñez 2005–) realizada en La Gomera en noviembre de 2018.

indiferencia del uso de mayúsculas y minúsculas) seguidas de números o de las palabras *muchos/pocos*.²⁴ El total de formas resultó de 944 en esTenTen y 14 en ProVaIn para *hubieron* y de 2 772 en esTenTen y 95 en ProVaIn para *habían*. Debido a la escasez de datos de la forma *hubieron*, los datos de ambas formas se presentan conjuntamente en el gráfico 8, donde se muestra el ratio entre formas concordadas y formas no concordadas en los contextos estudiados; en dicho gráfico puede observarse que, en aquellos casos en los que el número de ocurrencias sobrepasa las 15 ocurrencias en ambos corpus, ProVaIn presenta una mayor proporción de datos de concordancia.

En cuanto a las formas concordadas en primera persona, se realizó la búsqueda de la forma *habemos* en mayúsculas y minúsculas tanto en esTenTen como en ProVaIn,²⁵ lo que devolvió las cifras de 7 861 y 226 casos respectivamente²⁶. De nuevo, en el gráfico correspondiente (9) puede observarse que ProVaIn muestra una proporción más elevada del fenómeno cuando el número de ocurrencias supera los quince elementos.

²⁴ Con la expresión regular "((\\b[Hh][Aa]([Bb])([Vv])([Ii])\\xcd\\xed)[Aa][Nn]\\b)(\\b[Hh][Uu]([Bb])([Vv])([Ii][Ee][Rr][Oo][Nn]\\b))\\s(([Mm][Uu]([Cc][Hh])([Xx])[Oo][Ss])([Pp][Oo]([Cc][Kk])[Oo][Ss])([0123456789]))"

²⁵ Con la expresión regular "[Hh][Aa][Bb][Ee][Mm][Oo][Ss]".

²⁶ Debido a la búsqueda realizada, en el grueso de datos también se encuentran casos de *habemos* con valor posesivo (v.g. *habemos pantalones*), los cuales también representan el mantenimiento de un esquema (morfológico y semántico) antiguo y no estándar, pero distinto del que nos ocupa, y que precisaría su propio análisis. Una cala en 400 ejemplos nos permite ver que este uso es minoritario, siendo de en torno al 5% de las ocurrencias recogidas.

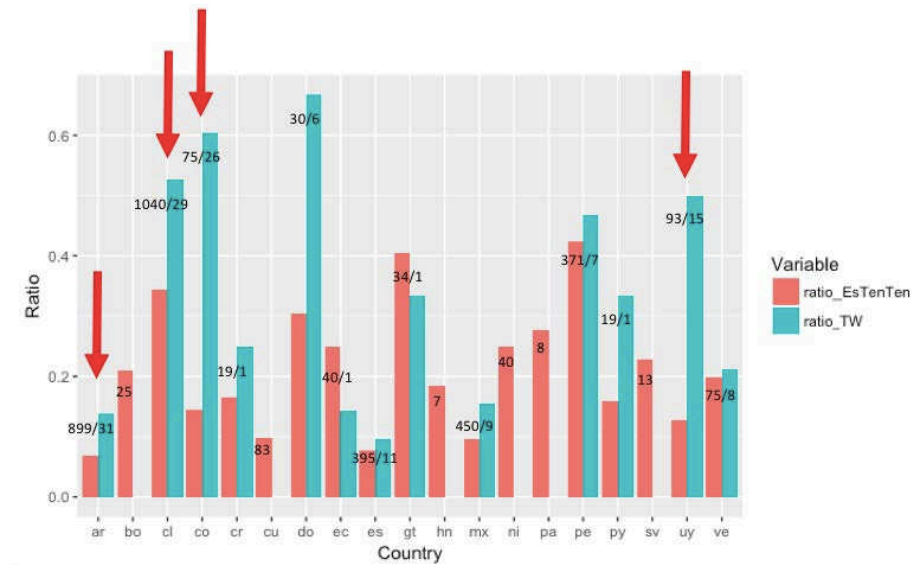


Gráfico 8. Ratio entre formas concordadas y no concordadas de {*habían/hubieron*} {*muchos/pocos/dígito*}. Los números absolutos de las formas concordadas aparecen sobre las barras.

País	% ETT	CI 95% ETT	% ProVaIn	CI 95% ProVaIn
Argentina	6.88%	6.44 – 7.31	13.72%	9.23 – 18.20
Chile	34.47%	32.77 – 36.17	52.73%	39.53 – 65.92
Colombia	14.42%	11.40 – 17.44	60.46%	45.85 – 75.08
Uruguay	12.79%	10.36 – 15.21	50%	32.10 – 67.89

Tabla 10. Porcentaje de datos con concordancia en los contextos estudiados.

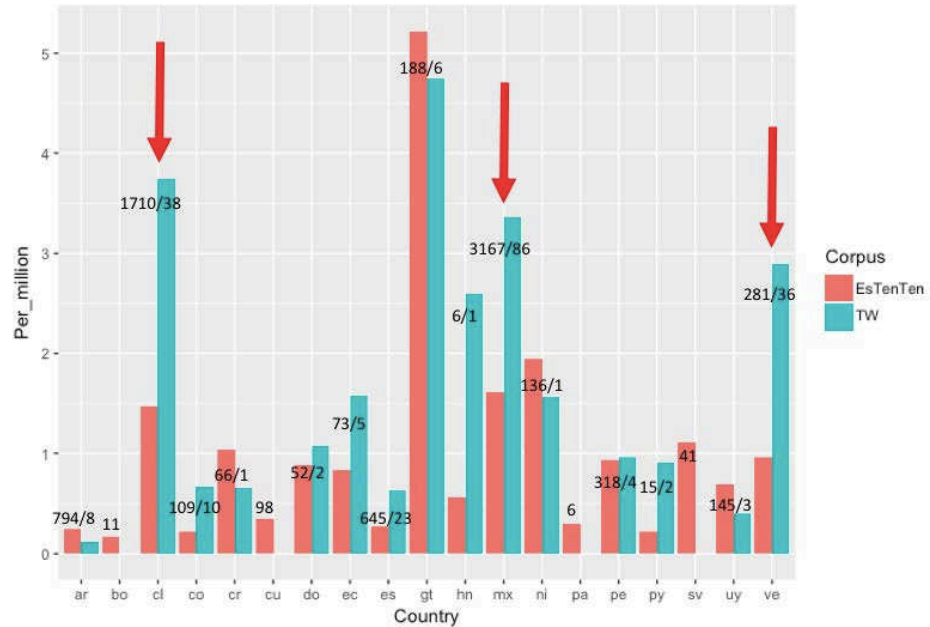


Gráfico 9. Frecuencia de *habemos* por millón de palabras

País	Por millón ETT	CI 95 % ETT	Por millón ProVaIn	CI 95 % ProVaIn
Chile	1.47	1.396 – 1.535	3.75	2.557 – 4.941
México	1.61	1.551 – 1.663	3.36	2.652 – 4.073
Venezuela	0.96	0.844 – 1.0678	2.89	1.943 – 3.829

Tabla 11. Frecuencia de *habemos* e intervalos de confianza por millón de palabras.

Algo que se puede observar en ambos gráficos es la gran extensión geográfica del fenómeno, que se encuentra en la mayoría de países hispanoamericanos con mayor o menor frecuencia. Si bien el número de ocurrencias es superior en esTenTen, en aquellos países en los que hay más de quince ejemplos por país (considerados necesarios para poder realizar generalizaciones estadísticas) parece que ProVaIn es proporcionalmente más favorable a la aparición del fenómeno (vid. Tablas 10 y 11).

Es importante no olvidar que los datos que tenemos no nos permiten sacar conclusiones sobre las diferencias entre pluralización para primera y tercera persona, ya que las búsquedas que hemos realizado no son de ningún modo equivalentes: mientras que se han podido registrar todas las ocurrencias de la forma *habemos* en su grafía estándar (y por tanto todos

los posibles usos concordados en primera persona dentro de los corpus utilizados), esto no es así para la tercera persona, ya que los contextos de búsqueda se han restringido para reducir al mínimo el número de usos estándar del plural.

Asimismo, sería necesario realizar un análisis más exhaustivo de las posibles lecturas en los casos de pluralización con *habemos* (v.g. su uso como auxiliar –*habemos venido*– frente a su uso existencial –*habemos cinco personas* (Castillo Lluch & Octavio de Toledo, 2016)), pero esto excede los objetivos de este trabajo. Sí nos es posible observar, sin embargo, que dentro de la tercera persona es efectivamente mucho más frecuente la forma *habían*.

En último lugar nos ocupamos de la forma pluralizada de *haber* impersonal en presente, cuya búsqueda supone un problema filológico de partida: ¿cómo se representa gráficamente? Se trata de una forma totalmente novedosa, con la peculiaridad morfológica de añadir un morfema –*n* tras un diptongo decreciente /ai/ y la peculiaridad ortográfica de que dicho diptongo se representa como *ay*, forma que no admite en principio la adición de otras consonantes a continuación. Por lo tanto esta forma no cuenta con un modelo de escritura claro, sino que existen múltiples posibilidades. Decidimos buscar las grafías *hayn*, *hain* y *haen*, además de sus contrapartidas sin hache: *ayn*, *ain*, *aen*.

Esta búsqueda en ProValn,²⁷ tras haber eliminado los casos de *ain't*,²⁸ contracción inglesa que aparece frecuentemente en títulos de canciones y, por tanto, en tuits en español, arrojó un total de 185 resultados. Todos estos casos resultaron ser nombres propios (*Ayn/Ain Rand* o *Ain el Muftella*, por ejemplo); erratas (*ain* por *sin*; *hayn* por *hayan*, *haen* por *hacen*, separaciones erróneas de palabras, etc.); la exclamación onomatopéyica *ain/ayn*, que expresa un suspiro, o, incluso, casos de la función lúdica del lenguaje, como el uso de *ain sorry* por *I'm sorry*. El único caso que podría tener una interpretación existencial es el siguiente, aunque también permite la interpretación de onomatopeya para indicar ‘suspiro’:

(18) Ain celos de amigas.??????????? <https://t.co/1mDO41GAQ4>“
(ID: 734968738043175000)

²⁷ Con la expresión regular “\b[Hh]?[Aa][YyIiEe][Nn]\b”.

²⁸ Con la expresión regular “\b[Aa][Ii][Nn].[Tt]\b”.

Los resultados de una búsqueda equivalente en esTenTen (en todos los países) son mucho más numerosos.²⁹ Sin embargo, suelen ser del mismo tipo que los encontrados en ProVaIn (nombres propios, onomatopeyas, erratas). Las grafías *aen* (1312 resultados) o *hain* (659 resultados) no ofrecen ningún resultado interpretable como existencial, mientras que las grafías *ain*, *ayn* y *haen* (18 048, 1 437, 192 resultados respectivamente) ofrecen cada una un resultado interpretable como existencial plural (véase (19)). El análisis de estas grafías, por lo tanto, ofreció un porcentaje de éxito muy bajo (solo 3 resultados “válidos” de 21 648, es decir, un 0.01 % de éxito), lo que supone un esfuerzo de anotación relativamente grande para obtener unos resultados bastante escasos.

(19)

- a) Lo veis insensatos?!?!?! La letra es profundísima!!! **Ain almas en pena**, y vosotros que queráis quitarle la razón a los gurús (España)
- b) donde ay arena que ay que levitarla con el baston **ayn tamien dos antorchas** que ya las e encendido i dos (España)
- c) peor hipócrita del mundo, porque detrás de él, **haen toda una verdadera caterva de malnacidos** que lo (Cuba)

La búsqueda de *hayn* resulta más productiva, sin embargo. De 334 resultados, 19 son casos interpretables como existenciales en plural (algunos ejemplos en 20) (un 5,7 % de los resultados obtenidos, que, si bien es un porcentaje de éxito sustancialmente mayor al de otras búsquedas, sigue ofreciendo un número muy bajo de casos para analizar). Sin embargo, documentamos también 16 casos con interpretación existencial en singular (21), lo que parece sugerir que se trata de erratas. Que los casos de *hayn* con objetos plurales y singulares presenten cifras

²⁹ No analizamos todos los casos manualmente. Para el caso de *ayn*, por ejemplo, eliminamos todos los casos que contenían la palabra *Rand* (con la expresión regular "[Rr]and"), para evitar el nombre propio. En el caso de *ain* (16496 resultados originalmente, eliminamos los casos totalmente en mayúsculas (*AIN*), por ser estas las siglas de una institución cubana y los casos seguidos de *ain't* (con las expresiones regulares "AIN", "\b[Aa][Ii][Nn].[Tt]\b" y "\b[Aa][Ii][Nn] >.[Tt]\b"). Para *aen* también eliminamos todos los casos en mayúsculas, también por ser siglas. En el caso de *hain* quitamos los casos precedidos de *el/del/al/un*, pues esta palabra funciona como sustantivo para designar un rito nativo de Tierra de Fuego.

tan similares nos hace pensar que los ejemplos en plural pueden también ser erratas y ser, por lo tanto, poco fiables.

(20)

- a) tal vez quì bajo la tierra **hayn espíritos** de muertos que matò Pinochet (lfcsiempre.blgoo.cl, Chile)
- b) Aunque **hayn jugadores** que merecen dignísima mencion (diariovoces.com.pe, Perú)

(21)

- a) Desde ayer no **hayn forma** de enviar correos (google.com.ni, Nicaragua)
- b) Aquí hay que tener cuidado porque **hayn un tipo de este pan** que no es libre de gluten (celiacosdemexico.org.mx, México)

La distribución por países (tabla 12) no nos da más confianza en la validez de estos datos, pues si solamente se trataran de erratas en el caso de la interpretación singular esperaríamos encontrar que los casos de plural estuvieran más concentrados geográficamente, mientras que las erratas deberían tener una distribución aleatoria. Encontramos, sin embargo, que los primeros aparecen en seis países y los segundos, en siete, cinco de los cuales son los mismos. Estos cinco países, además, están dentro de los 6 países con más volumen de palabras en esTenTen, por lo que son también aquellos países en los que esperamos encontrar más erratas (de cualquier tipo). Esta situación, por lo tanto, también nos hace dudar de la fiabilidad de los ejemplos de (19), que también son muy esporádicos.

País	PL	SG	Total
Argentina	8	5	13
Chile	2	2	4
México	2	4	6
Nicaragua	0	1	1
Panamá	0	1	1
Perú	3	1	4
Venezuela	2	0	2
España	3	2	5

Tabla 12. Distribución de los casos de *hayn* plurales y singulares por país en EsTenTen

En conclusión, el caso de la pluralización de *hay* nos muestra muy claramente una de las limitaciones de los grandes corpus de comunicación digital: la comunicación escrita poco planificada o controlada presenta un elevado número de erratas que, para fenómenos de muy baja frecuencia y cuya codificación escrita o es problemática (como el caso de *hay+n*) o se diferencia en solo una grafía de la forma estándar (por lo que la aparición de una letra “espuria” es relativamente sencilla), nos devuelve muchos resultados que exigen un laborioso preprocesado para resultar en un bajo número de casos, que, además, presentan una fiabilidad muy baja.

5. EL SUFIJO –ÉRRIMO

El sufijo elativo –*érrimo* es una alternativa no productiva a –*ísimo*, que, en principio, solo se adjunta a un pequeño número de bases cultas (como *paupérrimo* o *misérrimo*) (NGLE, 2009: 7.4n). Sin embargo, la NGLE (2009), también nota que “en el español coloquial de muchos países se usan con intención sarcástica o paródica *buenérrimo*, *elegantérrimo*, *guapérrimo*, *tristérrimo* y algunos otros derivados” (NGLE, 2009: 7.4n). Las Academias no concretan en qué países se encuentran estos usos sarcásticos o paródicos y en los estudios sobre el habla de culta de México (Madero Kondrat, 1983) o sobre el habla de Valparaíso (Peñailillo Fuentes, 2012) no se documentan. Esta última autora, de hecho, indica que –*érrimo* es poco frecuente en el habla coloquial.

Aparte de la restricción de las bases a las que se adjunta habitualmente, –*érrimo* también está sometido, según la NGLE (2009: 13.3a) a la restricción general que de no adjuntarse a adjetivos de grado extremo o elativos (*atroz*, *brutal*, *enorme*, *fabuloso* o *terrible*). Esta restricción, sin embargo, no está exenta de excepciones, como las propias Academias indican, y los casos de doble elación se documentan se han documentado con cierta frecuencia a lo largo de la historia (Pons Rodríguez, 2012).

En este caso nos interesamos, por lo tanto, por la posibilidad de encontrar en la comunicación mediada por ordenador formas “del español coloquial”, es decir, diafásicamente marcadas, como son los adjetivos no cultos intensificados por medio del sufijo –*érrimo*.

La búsqueda del superlativo –*érrimo* en ProValn se hizo con una expresión regular que extrajera todos los tuits que contuvieran una

palabra acabada en la secuencia {é/e}rrim{o/a}(s): es decir, tomando en cuenta todas las posibilidades flexivas del sufijo así como la posibilidad de que la palabra estuviera o no tildada (y que fuera indiferente al uso de mayúsculas y minúsculas).³⁰ Esto produjo un total de 525 tuits, de los cuales un total de 408 correspondían a palabras en las que la terminación *–érrimo* se adjunta a bases cultas (a saber, *acérrimo*, *celebérrimo*, *libérrimo*, *misérrimo*, *nigérrimo*, *paupérrimo*, *pulquérrimo*, *salubérrimo* y *ubérrimo*) y, por lo tanto, es estándar, y otros cinco fueron descartados por estar ubicados en Brasil (en algunos no podía discernirse qué lengua se estaba empleando). Una vez descartados estos, nos quedamos con 112 ejemplos de uso no estándar de dicho sufijo, con 34 adjetivos distintos: *guapo*, *bueno*, *enorme*, *grande*, *majo*, *mono*, *malo*, *brutal*, *tope*, *adorado*, *amigo*, *bruto*, *caluroso*, *chungo*, *cojonudo*, *crudo*, *cuqui*, *épico*, *filtrado*, *fresco*, *genial*, *gordo*, *hermoso*, *ilustre*, *lento*, *mucho*, *naco*, *puto*, *querido*, *recomendado rico*, *roto*, *sensual*, *viejo*. Como puede observarse en el gráfico 10, la distribución de estos adjetivos nos muestra que la mayoría de las ocurrencias corresponden solo a tres adjetivos: *guapo*, *bueno* y *enorme*. Esta distribución no nos indica necesariamente que *–érrimo* sea significativamente más frecuente con estos adjetivos (para lo que necesitaríamos conocer la frecuencia de otros recursos de intensificación con los mismos adjetivos), sino que es probablemente el resultado de la mayor frecuencia de algunos adjetivos frente a los demás.³¹

³⁰ La expresión regular concreta fue la siguiente:
 "([Ee]\xc9\xe9)[Rr][Rr][Ii][Mm][OoAa][Ss]?\\b"

³¹ Madero Kondrat (1983) también documenta que *–ísimo* aparece más frecuentemente con solo algunos adjetivos, como *bello*, *guapo*, *grande* y *mucho*.

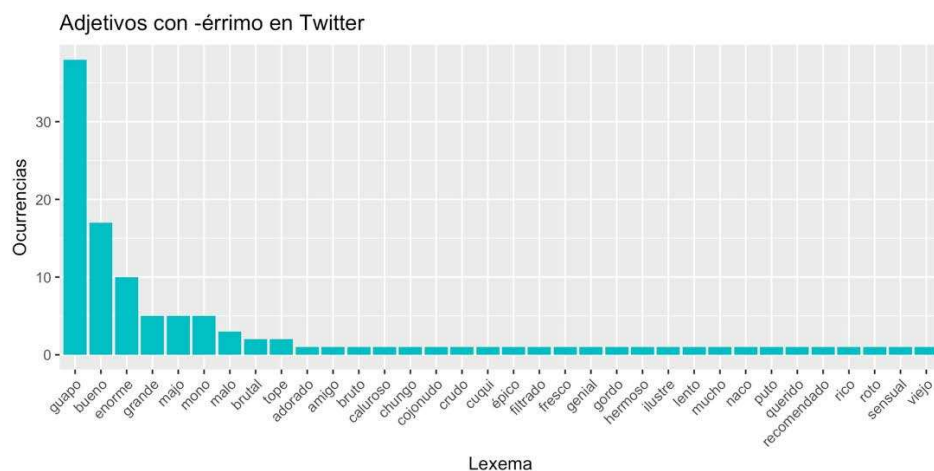


Grafico 10. Adjetivos que presentan *-érrimo* en el corpus ProVaIn

Todos estos casos se reparten en 4 países, aunque dos de ellos (Chile y la República Dominicana) solo atestiguan un ejemplo cada uno, hallándose la mayoría de los ejemplos en España, seguida de México (véase la tabla 13). Nótese, sin embargo, que estos números se igualan considerablemente cuando nos fijamos en la proporción de casos en las muestras de tuits de cada país: los intervalos de confianza entre la proporción de usos en España y México se solapa, por lo que no podemos afirmar que haya diferencias entre ambos países. Aquí se hace necesario comentar los diferentes pesos de cada país en el corpus ProVaIn (consultables en el Apéndice). Solo Argentina, España y México superan los diez millones de palabras en nuestro corpus. Estas diferencias hacen que sea difícil saber cómo de relevante es la ausencia de documentación en otros países, puesto que Chile y la República Dominicana solo documentan un ejemplo cada uno, imposibilitando el cálculo de estadísticas fiables. Sí resulta sugerente, sin embargo, la total ausencia de ejemplos en *-érrimo* en Argentina, que es el país más representado en nuestro corpus, con 36 679 535 tuits, es decir, un tamaño considerablemente mayor al de México o España, que documentan bastantes casos cada uno. Esto parecería ser indicativo de que los argentinos no emplean *-érrimo* como recurso coloquial en los contextos en que sí lo hacen los mexicanos y los españoles, lo que podría estar relacionado con la enorme productividad del prefijo elativo *re-* en la variedad Argentina (Kornfeld & Kuguel, 2013).

País	–érrimo	–érrimo por millón de palabras	Número de palabras por país
Chile	1 (0,9 %)	0,175	5 701 846
República Dominicana	1 (0,9 %)	0,941	1 062 227
España	71 (63,4 %)	3,192 (2,45-3,935)	22 240 135
México	39 (34,8 %)	2,752 (1,888-3,616)	14 171 764
Total	112 (100 %)	–	–

Tabla 13. Ocurrencias totales de –érrimo por país en el corpus ProVaIn (entre paréntesis, los intervalos de confianza al 95 %)

La búsqueda del superlativo –érrimo en esTenTen se hizo también buscando todas aquellas palabras acabadas en la secuencia {é/e}rrim{o/a}(s):³² es decir, buscando los mismos resultados que en ProVaIn. Esto produjo un total de 36 767 resultados, de los cuales solo 355 corresponden a usos no estándar del sufijo superlativo, con 88 bases distintas, a saber, *guapo*, *bueno*, *último*, *corrupto*, *paquete*, *mono*, *malo*, *queridérrimo*, *uribe*, *agudo*, *falsa*, *famoso*, *enorme*, *puto*, *culto*, *amigo*, *baboso*, *chico*, *crudo*, *dulce*, *fino*, *fresco*, *grande*, *hermoso*, *horrible*, *jodido*, *mínimo*, *mítico*, *naco*, *negro*, *perro*, *triste*, *jota*, *aburrido*, *amable*, *amado*, *atarde*, *bacán*, *blanco*, *bonito*, *brutal*, *campeón*, *caro*, *chafa*, *chingín*, *científico*, *cinéfilo*, *contundente*, *cool*, *correcto*, *dormido*, *elegante*, *esplendoroso*, *estimado*, *estupendo*, *friqui*, *fuerte*, *general*, *genial*, *gracioso*, *grave*, *harto*, *lento*, *limpio*, *listo*, *loco*, *majo*, *mala*, *malvado*, *mamá*, *minúsculo*, *mismo*, *patético*, *pedo*, *pendejo*, *perro*, *pinche*, *poco*, *rico*, *santo*, *sensual*, *simpático*, *simple*, *tonto*, *tremendo*, *útil*, *vago*, *viejo*. Como puede observarse en el gráfico 11, la distribución de estos adjetivos nos muestra que la mayoría de las ocurrencias corresponden solo a tres adjetivos: *guapo*, *bueno* y *último*.³³ Es notable la

³² La búsqueda en esTenTen de *érrimo produce automáticamente todas las posibilidades flexivas de la terminación, lo que no hace la búsqueda *errimo, que solo ofrece las palabras acabadas exactamente en esta terminación, por lo que debimos buscar las demás por separado. En el corpus de América lo hicimos buscando *errima, *errimos, *errimas, mientras que en el corpus de España buscamos *errim* (que ofrece unos resultados algo más amplios, pues la secuencia no está restringida a hallarse al final de la palabra).

³³ El caso de *último* es muy peculiar: la forma documentada es *ultérrimo*, aparentemente con el significado de ‘último’, según se deduce de los ejemplos. Este caso es particular porque toma de hecho una base culta (*ulter*), cuyo superlativo en latín era, precisamente, *ultimus*. Este caso, por lo tanto, no es un caso coloquial de –érrimo, aunque sí un caso subestándar.

frecuente en México y entre 2,9 y 8,4 en España que en Argentina. Parece claro, pues, que en país rioplatense el uso de *–érrimo* es muy inferior a lo atestiguable en España y México.

País	<i>–érrimo</i>	<i>–érrimo</i> por millón de palabras	Número de palabras por país
Argentina	35 (9,9 %)	0,0105 (0,007-0,014)	3 338 506 104
Chile	5 (1,4 %)	0,0043	1 167 093 083
Colombia	5 (1,4 %)	0,0101	496 187 321
Cuba	1 (0,3 %)	0,0036	279 681 643
República Dominicana	1 (0,3 %)	0,0170	58 862 734
España	118 (33,2 %)	0.0501 (0,041-0,059)	2 355 412 620
Honduras	2 (0,6 %)	0,1890	10 581 727
México	183 (51,5 %)	0,0929 (0,079-0,106)	1 970 160 304
Nicaragua	1 (0,3 %)	0,0143	69 972 296
Panamá	1 (0,3 %)	0,0476	21 000 533
Perú	2 (0,6 %)	0,0059	33 8027 632
El Salvador	1 (0,3 %)	0,0272	36 727 649

Tabla 14. Ocurrencias totales de *–érrimo* por país en esTenTen (entre paréntesis, los intervalos de confianza al 95 %)

Sin embargo, el gráfico 12 muestra con claridad que, cuando medimos frecuencias relativas (en este caso, ocurrencias de *–érrimo* por millón de palabras), ProVaIn ofrece resultados bastante superiores, lo que otra vez señala en la dirección de que la lengua de Twitter refleja un grado de inmediatez más elevado que la lengua registrada por el esTenTen, pues los usos coloquiales son más frecuentes en Twitter. De hecho, en España el uso de *–érrimo* en ProVaIn es entre 41,5 y 96 veces más frecuente que en esTenTen y en México, entre 17,8 y 45,8 veces más frecuente (compárense los intervalos de confianza de las tablas 13 y 14).

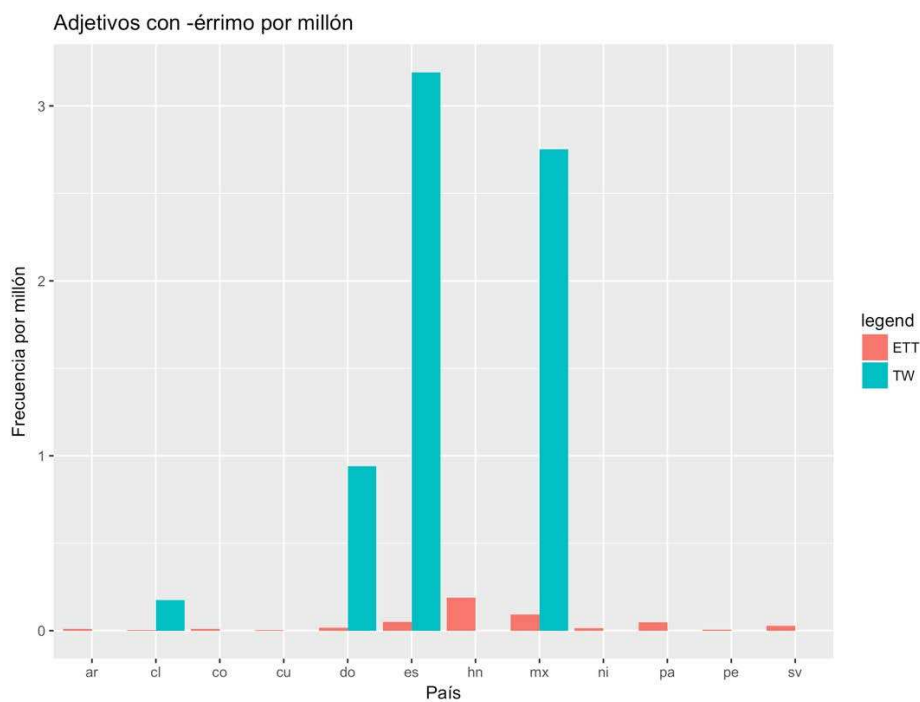


Gráfico 12. Frecuencia de *-érrimo* por millón de palabras por país y corpus

CONCLUSIONES

En este trabajo hemos comparado los resultados obtenidos en la investigación de distintos fenómenos de variación en esTenTen y el corpus del proyecto ProVaIn, formado por tuits. La finalidad del análisis era comprobar la utilidad de cada uno de estos corpus para el estudio de la variación lingüística, pues, aunque ambos contienen datos pertenecientes a la comunicación mediada por ordenador, sus características son muy distintas. Mientras que esTenTen es un corpus más general, que incluye todo tipo de páginas web, el corpus del proyecto de ProVaIn únicamente contiene tuits y, por lo tanto, representa un caso específico de interacciones digitales, más cercano a la inmediatez comunicativa que los textos que solemos encontrar en esTenTen. Otra importante diferencia entre ambos corpus es su tamaño, pues esTenTen es mucho mayor que el corpus ProVaIn (véase el apéndice). Por último, esTenTen resulta un corpus más sencillamente accesible, pues es un corpus ya compilado con su propia herramienta de búsqueda, mientras que todavía no existe un corpus de tuits disponible libremente y cada

investigador debe compilar el suyo —lo cual exige, además de tiempo, ciertas habilidades de programación (véase Estrada Arráez & De Benito Moreno, 2016a, b).

Las breves investigaciones que hemos llevado a cabo para realizar la comparación entre corpus nos permiten llegar a las siguientes conclusiones sobre ellos. Por un lado, la lengua documentada en ProVaIn es mucho más dialógica que la documentada en esTenTen, como muestran las frecuencias de los pronombres de 2ª pl. (sección 3.1). De hecho, esTenTen suele aparecer más cercano a corpus de lengua escrita no digital, como demuestran las comparaciones con el CREA y el CORPES XXI (sección 3.1 y 4 respectivamente). El hecho de presentar un contexto eminentemente dialógico favorece la inmediatez comunicativa, lo que explica que la incidencia de todos los fenómenos subestándar (ya estén marcados diatópicamente, diastráticamente o diafásicamente) aquí tratados sea mayor en términos relativos en el corpus ProVaIn que esTenTen. La mayor ventaja de esTenTen resulta ser su tamaño, que presenta consistentemente más ejemplos de los fenómenos buscados que el corpus ProVaIn en términos absolutos. La única diferencia es el caso de *ello* ante oraciones impersonales, en el que el pequeño ProVaIn presenta más ejemplos que esTenTen (y que el Corpus del Español Web y el CORPES XXI). Este es un caso muy interesante, pues se trata de una forma muy marcada tanto diatópica como diafásicamente y que, por tanto, no se atestigua en contextos de la distancia comunicativa. Twitter nos ofrece aquí una ventana única a la documentación de este fenómeno, a pesar de que su búsqueda sería más exigente en términos técnicos y de tiempo.

Otro punto de comparación interesante es el de la localización geográfica, que es mucho más precisa en ProVaIn, dando coordenadas o ubicaciones más concretas que en esTenTen o en otros macrocorpus de webs, en los que la localización se reduce a indicar el país. En estos últimos la información geográfica se obtiene a partir del dominio territorial de la URL, lo cual puede ser menos fiable (la movilidad física es menor que la movilidad virtual: parece más probable hacer un comentario en una página web con un dominio territorial que no corresponde con nuestro origen que estar o vivir en un país distinto del nuestro). Sin embargo, la comparación entre ambos corpus en el caso de la concordancia de *ustedes* + 2ª pl. (sección 3.2.) no muestra diferencias importantes en las frecuencias relativas entre países según el corpus, lo que nos hace pensar que en este aspecto se comportan de forma similar.

En términos absolutos, sin embargo, esTenTen introduce mucho más ruido, como vemos en el caso de *—érrimo* en Argentina (sección 6).

Por último, debemos hacer algunas observaciones por lo que respecta al estudio de la variación lingüística en cualquier corpus de comunicación mediada por ordenador. En primer lugar, por cercana a la inmediatez comunicativa que sea la lengua de algunos textos digitales, el habla siempre representa un grado mayor de inmediatez, pues la representación gráfica de fenómenos subestándar no es el mismo proceso que la producción de estos en el habla. En este sentido, el estudio de fenómenos fónicos a partir de corpus escritos supone una serie de problemas que hacen que este solo pueda ser complemento de estudios con datos orales si lo que queremos es una descripción adecuada del fenómeno. Por supuesto, el estudio de la representación gráfica de un fenómeno fónico puede tener interés en sí mismo (como bien demuestra Chariatte (2014)). Otro problema de —algunos géneros de— la escritura digital es que, debido a la profusión de contextos de la inmediatez comunicativa, caracterizados por un bajo grado de planificación, puede presentar un elevado número de erratas. Esto puede complicar el estudio de fenómenos fonéticos o morfológicos que implican solo pequeñas alteraciones en las formas estándar, como muestra el caso de *hayn* (sección 5). En este caso, la documentación de un número similar de formas con interpretación singular al de las formas con interpretación plural nos hace dudar necesariamente de la fiabilidad de todas las ocurrencias, por lo que el uso de comunicación escrita mediada por ordenador no parece un buen recurso para su estudio.

La mayor ventaja de los corpus de comunicación mediada por ordenador reside en que nos permiten acceder con relativa facilidad a grandísimas cantidades de lengua, incluso de lengua cercana a la inmediatez comunicativa, lo que supone que podemos encontrar una gran abundancia de ejemplos de fenómenos con baja frecuencia en otros corpus (especialmente los morfosintácticos). Sin embargo, nuestra investigación muestra que existen fenómenos, como el caso del *ello* no referencial (sección 5), con una frecuencia tan baja y una distribución tan restringida dentro del edificio variacional de la lengua que ni siquiera un macrocorpus como esTenTen, con casi once mil millones de palabras, nos permite documentarlo adecuadamente (o en absoluto, en este caso). A la inmensidad de la lengua el *big data* se le queda pequeño. Por último, conviene notar que el *big data* solo es tal para algunos lugares y algunas demografías: como se puede observar en el Apéndice, entre el 63 % (en

el corpus ProVaIn) y el 70 % (en esTenTen) de las palabras de cada corpus procede únicamente de tres países: Argentina, México y España. De entre los demás países, algunos de los corpus obtenidos son pequeños (lo que también puede explicar la escasa documentación de *ello* no referencial en la República Dominicana en ProVaIn, aunque más difícilmente en esTenTen, donde el subcorpus dominicano supera los 58 millones de palabras). Todo esto se debe, claro está, a una combinación de la cantidad de habitantes y el nivel de acceso a internet (correlacionado con el nivel socioeconómico) en los distintos países de habla hispana, y resulta en una distribución muy asimétrica entre países, que recuerda a la muy criticada representación de los distintos continentes en los corpus de referencia de la RAE.

APÉNDICE: EL VOLUMEN DE LOS CORPUS

País	EsTenTen: Número de palabras	ProVaIn: Número de palabras	EsTenTen: Porcentaje dentro del corpus	ProVaIn: Porcentaje dentro del Corpus
Argentina	3 338 506 104	36 679 535	30,4	33,7
Bolivia	64 369 251	141 976	0,6	0,1
Chile	1 167 093 083	5 701 846	10,6	5,2
Colombia	496 187 321	8 638 150	4,5	7,9
Costa Rica	63 446 946	842 176	0,6	0,8
Cuba	279 681 643	33 455 ³⁴	2,5	0,03
República Dominicana	58 862 734	1 062 227	0,5	1
Ecuador	87 518 018	1 852 695	0,8	1,7
El Salvador	36 727 649	346 574	0,3	0,3
Guatemala	35 995 321	701 162	0,3	0,6
Honduras	10 581 727	220 333	0,1	0,2
México	1 970 160 304	14 171 764	18	13
Nicaragua	69 972 296	363 992	0,6	0,3
Panamá	21 000 533	951 526	0,2	0,9
Paraguay	69 665 831	1 259 737	0,6	1,2
Perú	338 027 632	2 411 384	3,1	2,2

³⁴ La escasez de palabras en Cuba en el corpus ProVaIn no se debe a la baja penetración de internet en el país, sino a que, debido a esta, fue excluido expresamente de las coordenadas de búsqueda del proyecto ProVaIn. Los tuits encontrados se deben a que una pequeña área de Cuba quedaba dentro del polígono elegido para englobar los países centroamericanos.

Uruguay	211 291 691	4 184 374	1,9	3,8
Venezuela	293 912 478	7 104 498	2,7	6,5
España	2 355 412 620	22 240 135	21,5	20,4
Total	10 968 413 182	108 907 539	100	100

BIBLIOGRAFÍA

- Alba, O. (1999). Elisión de la /d/ intervocálica postónica en el español dominicano. En E. Forastieri Braschi, J. Cardona, H. López Morales y A. Morales de Walters (coords.), *Estudios de lingüística hispánica. Homenaje a María Vaquero* (3-22). San Juan, Puerto Rico: Universidad de Puerto Rico.
- Almeida, M. & Díaz Alayón, C. (1988), *El español de Canarias*. Santa Cruz de Tenerife: Litografía Romero.
- Bentivoglio, P. & Sedano, M. (1996). Tres casos de variación morfosintáctica en el español actual. *Lexis* 20(1), 165-184.
- Blas Arroyo, J. L. (2006). Hasta aquí hemos llega(d)o: ¿un caso de variación morfológica? Factores estructurales y estilísticos en el español de una comunidad bilingüe. *Southwest Journal of Linguistics* 25(2), 39-74.
- Cantamutto, L. y Vela Delfa, C. (2016). El discurso digital como objeto de estudio: de la descripción de interfaces a la de nición de propiedades. *Aposta: Revista de ciencias sociales* 69, 296-323.
- Castillo Lluch, M. & Octavio de Toledo y Huerta, Á. (2016). *Habemos muchos que hablamos español*: distribución e historia de la concordancia existencial en primera persona de plural. En C. de Benito Moreno y A. Octavio de Toledo y Huerta (eds.), *En torno a 'haber'. Construcciones, usos y variación desde el latín hasta la actualidad* (111-168). New York / Frankfurt: Peter Lang.
- Castillo-Trelles, C. (2007). La pluralización del verbo *haber* impersonal en el español yucateco. En J. Holmquist, A. Lorenzino, L. Sayahi (eds.), *Selected Proceedings of the 3th Workshop on Spanish Sociolinguistics* (74-84). Somerville, MA, Cascadilla Proceedings Project.
- Catalán, D. (1964). El español en Canarias. *Presente y futuro de la lengua española*. Madrid: OFINES I, 239-280.
- Chariatte, N. (2014). 'Facebook Style': The use of non-standard features in virtual speech conditioned by the medium Facebook. En: J. Brumme y S. Falbe, *The Spoken Language in a Multimodal Context* (93-117). Berlin: Frank & Timme.
- Claes, J. (2014), *The pluralization of presentational haber in Caribbean Spanish. A Study in Cognitive Construction Grammar and Comparative Sociolinguistics* (tesis doctoral). Universidad de Amberes. Amberes.

- Claes, J. (2016). La pluralización de *haber* presentacional en el español de la Habana: interacción entre las restricciones cognitivas. En C. De Benito Moreno y A. Octavio de Toledo (eds.), *En torno a haber. Construcciones, usos y variación desde el latín hasta la actualidad* (169-190). New York / Frankfurt: Peter Lang.
- Company Company, C. (1998). The interplay between form and meaning in language change. Grammaticalization of cannibalistic datives in Spanish. *Studies in Language* 22(3), 539-565.
- CORPES XXI = Real Academia Española: *Corpus del Español del Siglo XXI (CORPES)*. Recuperado de <http://www.rae.es>.
- Corpus del Español = Davies, M. (2016). *El corpus del español*. Recuperado de <http://www.corpusdelespanol.org/>
- Corr, A. (2016). El cambio lingüístico y la interfaz sintaxis-pragmática: evidencias de las lenguas ibero-románicas. *Cuadernos de lingüística* 3(1), 73-118.
- CREA = Real Academia Española: *Corpus de Referencia del Español Actual*. Recuperado de <http://www.rae.es>
- esTenTen = SketchEngine: *esTenTen*. Recuperado de <https://auth.sketchengine.eu/>
- Estrada Arráez, A. (2017). *The loss of intervocalic /d/ in the Iberian Peninsula* (tesis doctoral inédita). Universidad de Friburgo de Brisgovia. Friburgo de Brisgovia.
- De Benito Moreno, C. y Estrada Arráez, A. (2016a). Introducción. Si Alvar levantara la cabeza: cómo tratar con informantes virtuales. *Revista Internacional de Lingüística Iberoamericana* 14 (2)/28, 7-17.
- Estrada Arráez, A. y de Benito Moreno, C. (2016b). Variación en las redes sociales: datos twiectales. *Revista Internacional de Lingüística Iberoamericana* 14 (2)/28, 77-111.
- Fernández-Ordóñez, I. (2005-). *Corpus Oral y Sonoro del Español Rural*. Recuperado de <http://www.corpusrural.es/>
- Fontanella de Weinberg, M. B. (1992). Variación sincrónica y diacrónica de las construcciones con *haber* en el español americano. *Boletín de Filología de la Universidad de Chile* 33, 35-46.
- Fontanella de Weinberg, M. B. (1999). Sistemas pronominales de tratamiento usados en el mundo hispánico. En Bosque, I. y Demonte, V., *Gramática de la Lengua Española* (1399-1425). Madrid: Espasa.
- Heap, D. (2003). Sentencias ‘invertidas’ de clíticos: un cambio (¿?) en tiempo real. En J. L. Girón Alconchel y J. J. Bustos Tovar (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua Española* (785-798). Madrid: Arco / Libros.
- Henríquez Ureña, P. (1939). Ello. *Revista de Filología Hispánica* 1(3), 209-229.
- Hinzelin, M. O. & Kaiser, G. A. (2007). El pronombre *ello* en el léxico del español dominicano. En W. Mihatsch y M. Sokol (eds.), *Language contact and language*

- change in the Caribbean and beyond* (171-188). Frankfurt am Main: Lang.
Recuperado de: <https://d-nb.info/1093432330/34>.
- Hinzelin, M. O. (2010). Catalan and Spanish neuter pronouns in impersonal constructions: diachronic and dialectal variation. *Dialectologia* 4, 23-45.
- Kilgarriff, A. & Renau, I. (2013), esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia - Social and Behavioral Sciences* 95, 12-19.
Recuperado de: <http://doi.org/10.1016/j.sbspro.2013.10.617>.
- Koch, P. & Oesterreicher, W. (2007), *Lengua hablada en la Romania: español, francés, italiano*. Madrid: Gredos.
- Kornfeld, L. y Kuguel, I. (2013). Un afijo re loco (Notas sobre re). En A. Di Tullio (ed.), *El español de Argentina: estudios gramaticales* (13-33). Buenos Aires: Eudeba.
- Lara Bermejo, V. (2015), *Los tratamientos de 2pl en Andalucía occidental y Portugal. Estudio geo- y sociolingüístico de un proceso de gramaticalización* (tesis doctoral). Universidad Autónoma de Madrid. Madrid. Recuperado de: <http://hdl.handle.net/10486/667216>.
- Madero Kondrat, M. (1983). La gradación del adjetivo en el habla culta de la ciudad de México. *Anuario de Letras. Lingüística y Filología* 21, 71-118.
- Menéndez Pidal, R. (1962), *Manual de gramática histórica española*, Madrid: Espasa Calpe.
- Molina Martos, I. & Paredes García, F. (2014). Sociolingüística de la elisión de la dental -/d/- en Madrid (distrito de Salamanca). *Cuadernos de Lingüística* 2, 55-114.
Recuperado de: http://cuadernoslinguistica.colmex.mx/images/PDF/Volumen2/Molina_Paredes.pdf
- Morgan, Terrel A. & Schwenner, S. A. (2016). *Vosotros, ustedes*, and the myth of the symmetrical Castilian pronouns system. En A. Ciza, L. Czerwionka y D. Olson (eds.), *Inquiries in Hispanic Linguistics. From theory to empirical evidence* (263-279). Amsterdam / Philadelphia: John Benjamins.
- Moya Corral, J. A. & García Wiedemann, E. J. (2010). La elisión de /d/ intervocálica en el español culto de Granada: factores lingüísticos. *Pragmalingüística* 17, 92-123.
Recuperado de: <http://revistas.uca.es/index.php/pragma/article/view/166/169>
- Muñoz Pérez, C. (2014). Dominican *ello* as a non-deleted null expletive. *Borealis. An International Journal of Hispanic Linguistics* 3, 155-161. Recuperado de: <http://dx.doi.org/10.7557/1.3.1.2945>
- NGLE (2009) = RAE / ASELE (2009), *Nueva gramática de la lengua española*. 2 vols. Madrid: Espasa.
- Octavio de Toledo & Huerta, Á. (2016). Sin CORDE pero con red: *algotras* fuentes de datos. *Revista Internacional de Lingüística Iberoamericana* 14 (2)/28, 19-48.

- Pato Maldonado, E. (2016). La pluralización de *haber* en español peninsular. En C. de Benito Moreno y Á. Octavio de Toledo y Huerta (eds.), *En torno a 'haber'. Construcciones, usos y variación desde el latín hasta la actualidad* (357-391). New York / Frankfurt: Peter Lang.
- Pavalanathan, U. & Eisenstein, J. (2015). Confounds and consequences in geotagged twitter data. En Ll. Màrquez, C. Callison-Burch y J. Su, J. (eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2138-2148). Lisbon, Portugal: arXiv.
- Peñailillo Fuentes, M. I. (2012), *Estudio sociolingüístico sobre el uso del superlativo en la región de Valparaíso, Chile* (tesis doctoral). Universidad de Valladolid.
- Pérez Martín, A. M. (2007). Pluralización de *había* en el habla de *El Hierro*: datos cuantitativos. *Revista de Filología* 25, 505-513.
- Pons Rodríguez, L. (2012). La doble graduación muy -ísimo en la historia del español y su cambio variacional. En E. Pato y J. Rodríguez Molina (eds.), *Estudios de filología y lingüística españolas* (135-166). New York / Frankfurt: Peter Lang.
- Quintanilla-Aguilar, J. R. A. (2009), *La (des)pluralización del verbo haber existencial en el español salvadoreño: ¿un cambio en progreso?* (tesis doctoral). University of Florida. Miami. Recuperado de: http://digitalcommons.butler.edu/cgi/viewcontent.cgi?article=1412&context=facsch_papers.
- Samper Padilla, J. A. & Hernández Cabrera, C. E. (2012). En torno a los usos personales de *haber* en el español de Las Palmas de Gran Canaria. En T. Jiménez Juliá, B. López Meirama, V. Vázquez Rozas y A. Veiga (eds.), *Cum corde et in nova grammatica: estudios ofrecidos a Guillermo Rojo* (743-754). Santiago de Compostela: Universidade de Santiago de Compostela.
- Toribio, A. J. (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua* 10, 315-341.